THIS WEEK

EDITORIALS

WORLD VIEW Policy-makers must engage with the reality of geoengineering **p.159**

WEIGHT LOSS A true measure of the mass of extinct beasts **p.160**

BLUE DANUBE The cultural flow of the river 40,000 years ago **p.161**

Reply to all

Scientists discussing their work through written media, including e-mail, should be aware that they could at any time be asked to reveal their conversations.

"You are commanded to produce … any and all documents, data, and/or communications." Towards the end of last year, those orders appeared in a subpoena that landed at the Woods Hole Oceanographic Institution in Massachusetts. The energy firm BP demanded that Woods Hole produce e-mails and other documents related to its research on the 2010 Deepwater Horizon oil spill in the Gulf of Mexico. Woods Hole fought the sweeping request, but a US district court has now forced researchers at the institute to surrender thousands of e-mails. That decision has disturbing implications for science in the United States, although the situation is perhaps not as dire as some have warned.

The demand for the e-mails emerged from a huge lawsuit, in which BP is being sued by the US government and others affected by the oil spill. As part of that suit, the company faces fines of up to US$4,300 per barrel of oil spilled, which could amount to more than $17 billion if the court sides with government estimates for the size of the spill. BP argues that these estimates — much higher than the company's own — rely heavily on research conducted by Woods Hole. And it claims that aspects of the work involve "puzzling, apparently arbitrary, suspiciously offsetting, and highly significant decisions by Woods Hole researchers".

Woods Hole is not a party to the lawsuit, but it did have a contract with the government during the first month of the spill to measure how much oil and gas was spewing from the broken well. After that, researchers from the institute worked to refine their analyses and publish papers on the spill in academic journals. One Woods Hole scientist also participated in a governmental panel asked to estimate the size of the spill.

To defend itself in court, BP contends that it needs more than just the raw data collected by Woods Hole and a description of the methods used by its researchers. Woods Hole turned over much of the data and analysis tools to BP on its original request, but it fought against surrendering confidential academic communications, arguing that those have been protected in the past by courts, which have recognized the importance of 'scholastic privilege'. Indeed, the judge who ruled over Woods Hole's arguments recognized the principle of scholastic privilege regarding confidentiality, but only up to a point. She found that BP had a compelling need for the Woods Hole e-mails and other communications for the period before the government-led group issued its report in March last year. The e-mails handed to BP could be made public if the case goes to trial.

But the judge did not allow BP access to communications after that time, when the Woods Hole researchers were preparing academic publications. "If BP gains access to the analysis documents for these article [*sic*], that could hamper future research efforts," she said.

The decision pleased neither BP nor Woods Hole. BP appealed, seeking all the documents it requested, but was unsuccessful. Woods Hole contends that surrendering the e-mails will have a chilling effect on science, by forcing researchers to avoid topics that could become subjects of litigation. Is the institute right? It is hard to judge the scale of the potential damage the decision could cause. Even well before the subpoena, some researchers declined to study the spill, in part because of the legal risk, so the decision may not make much difference on that front. Researchers who work in other contentious fields, such as animal experimentation, are not generally scared off their work, even by threats of physical harm.

"It is hard to judge the scale of the potential damage the decision could cause."

The decision on the Woods Hole e-mails should nonetheless serve as a warning. Researchers who choose to study topics that have multiple and conflicting implications for powerful special interests must go in with their eyes open and be aware of the risks. Not only have Woods Hole researchers had to surrender their e-mails, they have also spent hundreds of hours complying with the subpoena. And the case has drained considerable funds from the non-profit institution through lawyers' fees and lost staff time.

The decision provides yet another reminder that scientists should not regard e-mails as being private. Whether by theft or by court order, such communications can be taken away. Think before you press Send. ■

ANNOUNCEMENT

Calling Nordic mentors

Every year, *Nature* awards prizes to senior researchers in the natural sciences for sustained and outstanding scientific mentoring of young researchers. There are two prizes, each worth €10,000 (US$12,500): one for lifetime achievement, one for mid-career achievement.

The competitions are organized on a country or regional basis. We shall give this year's accolades to researchers who live in the Nordic countries: Sweden, Norway, Denmark, Finland, Iceland and Greenland.

Nominations are now open, with a deadline of Monday 13 August. Each applicant should be nominated by five people who have been mentored by her or him over different periods. The forms for nominations, and other essential details about the competition, including the list of judges, can be found at go.nature.com/bxzc9y.

WORLD VIEW *A personal take on events*

Researchers can't regulate climate engineering alone

Political interests, not scientists or inventors, will be the biggest influence on technologies to counter climate change, says **Jason Blackstock**.

Scientists are developing geoengineering technologies. But whether these methods eventually succeed in countering climate change, and whether they will be embraced by the wider population, concerns more than scientists alone. That is why, in the wake of the cancellation last month of the Stratospheric Particle Injection for Climate Engineering (SPICE) field trial and the attendant publicity, it is important that the correct lessons are drawn by scientists, funders, regulators and politicians alike.

In an Editorial about SPICE, this publication pointed out that geoengineers must "grasp the nettle of regulation and oversight" (see *Nature* **485,** 415; 2012). It added that community-proposed guidelines such as the Oxford Principles need practical implementation frameworks to make an impact. Both issues do demand urgent attention, but there is a more important consideration: something that geoengineers coming together to "draft detailed, practical actions", as the Editorial recommended, simply cannot address. Geoengineering will alter the geopolitics of climate change and this cannot be ignored by climate policy-makers.

SPICE turned the focus of the geoengineering debate onto patenting and concerns that personal or corporate ownership could take precedence over global public interest. But patents are far from the biggest issue for techniques such as solar geoengineering (blocking the Sun's radiation, as investigated by the SPICE project).

The cost of injecting enough aerosols into the stratosphere to counter the warming projected for even high-emission scenarios is estimated to be only about US$1 billion per year. That amounts to less than $0.01 per year to compensate for each tonne of carbon dioxide emitted. And most of the necessary technology exists already, such as high-altitude aircraft (of which only a couple of dozen would be needed). Inventors of solar-geoengineering methods might try to charge large sums to license patents on their ideas, but this will not create a lucrative new industry for large corporations.

That said, existing industries, especially agriculture and energy, will have a strong interest in whether and when solar geoengineering is used, given that these methods could have sweeping effects on climate, both desirable and undesirable. But industry influence will occur much more through political lobbying than through patent ownership.

At present, the power in climate negotiations is squarely in the hands of the major carbon-emitting nations. Technically, a coalition of vulnerable nations — say, low-lying island states threatened by rising seas — might be able to muster $1 billion per year to reverse global warming, but the potential game-changer when it comes to geoengineering is not technical assessments, but geopolitical calculations. We scientists know this. We have consulted with civil society, the private sector and government officials through the 2010 Asilomar Conference in California — which worked on policy recommendations for geoengineering — and the ongoing Solar Radiation Management Governance Initiative launched by the Royal Society in London. Both have indicated the need for action beyond geoengineering researchers. But if assembling these groups has not yet generated the necessary 'practical actions', then what will?

GEOENGINEERING WILL ALTER THE GEOPOLITICS OF CLIMATE CHANGE AND THIS CANNOT BE IGNORED.

➲ NATURE.COM
Discuss this article online at:
go.nature.com/ws6bxs

Geoengineering researchers can experiment with ways to translate guidelines into a practical regulatory framework, beginning with a voluntary registry of ongoing and planned research. But responsibility also rests with funders. More research is needed to ensure that we understand the abilities and limits of solar-geoengineering technology before it is deployed in political desperation. It would be a tragedy if the negative publicity around SPICE scared funds away from this work. Funders can also help to establish norms of international cooperation. First steps would include creating incentives for collaboration on research, and sharing lessons on how to responsibly oversee and monitor controversial projects.

Ultimately, climate policy-makers need to get involved. Current international institutions are simply not prepared for geoengineering. Discussing solar geoengineering at the United Nations climate talks now would only disrupt progress towards the already ambitious goal of agreeing a global mitigation and adaptation strategy by 2015. And the Convention on Biological Diversity, the only international body yet to tackle the issue of solar geoengineering directly, has neither the mandate nor sufficient political clout to broker a geopolitical agreement.

Climate negotiators and political leaders need to develop strategies to fill the governance gap. They also need to consider the signals that domestic funding of geoengineering research sends about future climate-policy intentions. Failure to come to grips with these issues could lead to problems if events such as geoengineering field experiments outpace political preparations.

Greater political engagement may be uncomfortable for climate scientists. Climate science has fought hard to resist agendas that force particular research or conclusions. But such agendas — both corporate and national — do influence climate policy, and geoengineering technologies could magnify that influence. Politics can't be avoided, and ignoring it is dangerous for all of us. ■

Jason Blackstock *is a visiting fellow with the Institute for Science, Innovation and Society at the University of Oxford, UK.*
e-mail: jason.blackstock@insis.ox.ac.uk

RESEARCH HIGHLIGHTS

Selections from the scientific literature

CULTURE

Religion as cuckoldry defence

How can patriarchal societies be sure that the men who inherit their fathers' goods are actually their fathers' sons? Religions help men to be more confident in their paternity by setting standards for female sexual behaviour.

Beverly Strassmann at the University of Michigan in Ann Arbor and her team studied genetic data from 1,706 father–son pairs among the Dogon of Mali. Although many Dogon are Christian or Muslim, the adherents of their indigenous religion had the lowest rate of male births due to cuckoldry: 1.3% versus 2.9% for the other faiths. The researchers chalk this up to the traditional huts where menstruating women must spend their nights. In a society in which women are often pregnant or unable to conceive owing to intensive breast-feeding, a sojourn in the hut informs a woman's husband and his kin that she has become fertile again. The family can then increase their vigilance towards her.

Proc. Natl Acad. Sci. USA **http://dx.doi.org/10.1073/pnas.1110442109 (2012)**

BIOMIMETICS

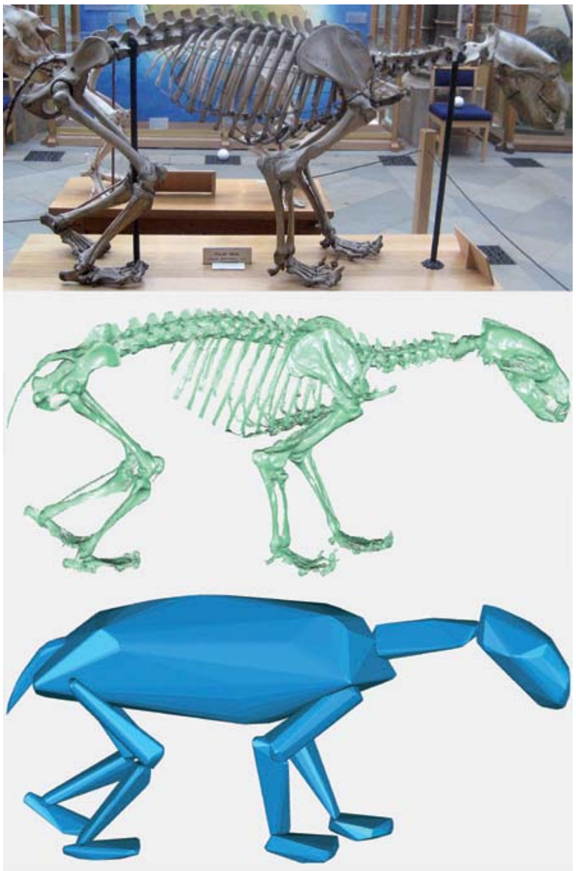
Synthetic silk inspired by insect

The green lacewing (*Mallada signata*), an insect used in biological pest control, has inspired the creation of an artificial silk.

To protect its progeny from predators, the female lacewing suspends its eggs from the underside of a leaf using a thin silk stalk of remarkable tensile strength. To forge a similar stalk, Thomas Scheibel and Felix Bauer at the University of Bayreuth in Germany developed a synthetic protein based on a repeating amino-acid pattern found in one of the proteins in lacewing silk. The researchers introduced a synthetic gene that codes for the protein into *Escherichia coli* bacteria.

Filaments of the bacteria-produced silk proved to be almost as strong as natural lacewing silk, except at high humidity, and more elastic.

Angew. Chem. Int. Edn. **http://dx.doi.org/10.1002/anie.201200591 (2012)**



W. SELLERS

DINOSAURS

Weighing extinct animals

Knowing an extinct animal's mass is crucial for estimating its physiological traits but the standard technique — in which a model of the animal is made and its mass then calculated from its density — has been criticized for being too subjective.

Bill Sellers at the University of Manchester, UK, and his team laser-scanned the skeletons of 14 large mammals, including the bison and the elephant. The researchers used the scans to mathematically derive 'convex hulls' of the bones — the minimum volume that encloses a set of points, akin to gift-wrapping a teapot — and converted these volumes into estimates of mass. These were then compared to the known values.

The method consistently underestimated true body mass by 21%. Thus, using this method and then adding 21% should provide more accurate predictions. When the team used the corrected technique on *Giraffatitan brancai*, one of the largest dinosaur skeletons in the world, the beast clocked in at 23,200 kilograms, similar to recent volumetric estimates.

Biol. Lett. **http://dx.doi.org/10.1098/rsbl.2012.0263 (2012)**

ZOOLOGY

Live birth without a placenta

Some sharks and rays give birth to live young but lack placentas or umbilical cords. So how do their embryos obtain oxygen?

Taketeru Tomita of the Hokkaido University Museum in Hakodate, Japan, and his team used ultrasound to study a pregnant manta ray (*Manta alfredi*) to find out. They found that the embryo rhythmically raises and lowers its jaw, pumping uterine fluid into its mouth and through holes behind its eyes called spiracles. After birth, the spiracles shrink as the animal becomes able to force water over its gills by swimming.

Biol. Lett. **http://dx.doi.org/10.1098/rsbl.2012.0288 (2012)**

DRUG DISCOVERY

Melanoma pathway targeted

A drug that targets the RAS signalling cascade — which is thought to drive tumour growth in 20% of cancers — improves survival in some patients with advanced melanoma.

The drug, called trametinib, inhibits MEK, a protein that amplifies cancer-promoting signals. Keith Flaherty at the Massachusetts General Hospital in Boston and his team tested trametinib in melanoma patients with a mutated version of a RAS-pathway protein BRAF, which is mutated in 50% of advanced melanoma patients. A total of 322 patients were randomly assigned to receive either trametinib or standard chemotherapy.

Twenty-two percent of patients responded to trametinib, which delayed disease progression by 3 months longer than chemotherapy alone.

N. Engl. J. Med. **http://dx.doi.org/10.1056/nejmoa1203421 (2012)**

ARCHAEOLOGY

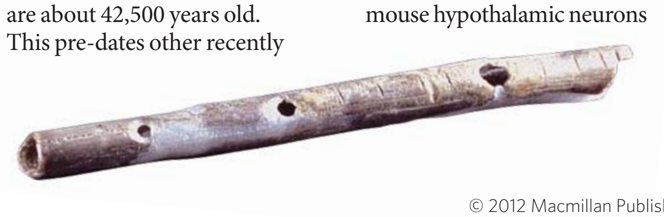
Cultural wellspring

Certain forms of art and music might have emerged among ancient humans living in southern Germany and then spread through Europe along the Danube River.

The Aurignacian culture, which is characterized by innovations such as figurative art and musical instruments (**mammoth-ivory flute pictured**), originated in Europe more than 40,000 years ago. However, imprecision in carbon dating has made it difficult to determine precisely where the culture arose. Thomas Higham at the University of Oxford, UK, Nicholas Conard at the University of Tübingen, Germany, and their team used improved sample preparation techniques for carbon dating to study Aurignacian remains from Geißenklösterle cave in southern Germany's Swabian Jura region.

The analysis revealed that the artefacts at Geißenklösterle are about 42,500 years old. This pre-dates other recently dated Aurignacian sites in Europe.

J. Hum. Evol. **http://dx.doi.org/10.1016/j.jhevol.2012.03.003 (2012)**



UNIVERSITY OF TÜBINGEN

EXTRASOLAR PLANETS

A new world is born

Astronomers have identified a planet in the making.

The variable brightness of GM Cephei — a 4-million-year-old star 870 parsecs away — has been noted for decades. But whether that variability is caused by bursts of light from the star or by an object in the disk of dust that surrounds it has been a matter of debate.

Wen-Ping Chen at the National Central University in Jhongli, Taiwan, and his team argue for the latter. Using telescopes positioned around the world, the researchers monitored GM Cephei from 2009 to 2011 and found that the brightness of the star dipped each year for about a month. The researchers interpret this variability as evidence of a young planet: a clump of dust, roughly the mass of an asteroid, which is orbiting GM Cephei.

Astrophys. J. 751: **118 (2012)**

METABOLISM

A pathway for feeding control

The activation of a group of neurons in a region of the brain called the hypothalamus boosts food intake, but what controls this activation has been unclear. Researchers have now identified a receptor in these neurons that is targeted by a protein called FoxO1 to promote feeding.

The FoxO1 protein is involved in regulating the signalling of key hormones that suppress appetite. Domenico Accili at Columbia University in New York and his team deleted FoxO1 from mouse hypothalamic neurons and found that the animals were leaner, ate less and had improved glucose and fat metabolism compared with normal mice. The researchers pinpointed the receptor Gpr17 as mediating the effects of FoxO1.

Targeting this pathway could be a strategy for anti-obesity drugs, the authors say.

Cell 149, **1314–1326 (2012)**

COMMUNITY CHOICE

The most viewed papers in science

EVOLUTION

Domesticated apes

HIGHLY READ on journals.elsevier.com/animal-behaviour March–May

Selection against aggression seems to have occurred naturally in some animals and to have led to traits similar to those seen in domesticated animals.

The bonobo (*Pan paniscus*) is less aggressive and more sociable than its sister species, the chimpanzee (*Pan troglodytes*). Brian Hare at Duke University in Durham, North Carolina, and his team propose that these differences, along with a suite of physical variations — bonobos have smaller childlike heads, plus paler lips and tails, than chimps — are a linked set of traits that parallel those seen in domesticated animals such as dogs and guineapigs. Bonobos are, the authors argue, a self-domesticated ape.

Selection pressures that might have favoured bonobo self-domestication include the acquisition of bigger territories, reducing competition for food and the emergence of coalitions of females that enforced the peace.

Domestication, far from being a human invention, may also occur spontaneously in nature, they say.

Anim. Behav. 83, **573–585 (2012)**

GEOPHYSICS

Ship GPS could flag tsunamis

Commercial ships on Earth's oceans could provide a cheap and easy way to track propagating tsunamis.

Current warning systems rely on sparse, expensive buoys and underwater sensors that track a wave once it has been triggered by an earthquake. James Foster at the University of Hawaii at Manoa in Honolulu and his team examined Global Positioning System (GPS) data from a research vessel (**pictured**) that was heading from Hawaii to Guam in February 2010 when

SOEST/UNIVERSITY OF HAWAII

an underwater earthquake occurred off the coast of Chile. Filtering the data on the basis of tsunami models allowed the researchers to differentiate between choppy seas and changes in ocean surface height due to a tsunami. The team detected a roughly 10-centimetre rise as the tsunami passed, and estimated wave speed and arrival time.

Although data from a single ship could be prone to false positives, recruiting thousands of ships could overcome this problem, the authors suggest.

Geophys. Res. Lett. **http://dx.doi.org/10.1029/2012GL051367 (2012)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Nuclear restarts

Two of Japan's shuttered nuclear reactors may soon reopen, having received a safety thumbs-up from a scientific panel appointed by regional government. The reactors, at the Ohi plant in Fukui prefecture, had already passed safety tests in January. But public protests (and some nuclear analysts) opposed their restart. On 8 June, Prime Minister Yoshihiko Noda appeared on public television asking for his nation's support to reopen the nuclear reactors, and on 10 June, a 12-member scientific panel appointed by the Fukui prefecture's governor reported that the Ohi plant could be operated safely.

X-ray mission axed

NASA has cancelled a small astrophysics mission that would have studied the polarized X-rays streaming from black holes and neutron stars. The Gravity and Extreme Magnetism Small Explorer (GEMS) had been selected for funding in 2009 and was scheduled to launch in 2014, but was estimated to be running 20–30% over its US$119-million budget by the time it was axed on 7 June. See go.nature.com/bit1rn for more.

Altering embryos

Reproductive procedures that would spare children from inheriting mitochondrial diseases received approval from an influential UK bioethics body on 12 June. The technique involves transferring genetic material from an egg cell containing faulty mitochondria to an egg from another woman that has been stripped of its nucleus (see *Nature* **481,** 419; 2012). The London-based Nuffield Council on Bioethics found that these procedures are ethically acceptable, if proven safe and effective by additional research currently under way, and if conducted as clinical trials. The UK government begins a public consultation in September, a first step to legalizing the process. See go.nature.com/fpn5bo for more.

J. FREUND/NATUREPL.COM

Great Barrier Reef protection row

State and federal governments in Australia are clashing over the approval of industrial developments that could damage the Great Barrier Reef. On 5 June, environment minister Tony Burke delayed the approvals process for a multibillion-dollar coalmining project in Queensland, calling the state's operation "shambolic". Three days earlier, a report from the United Nations Educational, Scientific and Cultural Organization (UNESCO) had said that the reef was threatened by the rapid increase in coastal development. UNESCO's Heritage Committee is meeting from 24 June to consider the report's findings — which stopped short of recommending a potentially embarrassing 'in danger' designation.

Environment gloom

Only four of the 90 most important internationally agreed environmental goals have seen "significant progress", the United Nations Environment Programme warned on 6 June. Its fifth Global Environment Outlook is the UN's main status report on the health of the world's ecosystems, and was released two weeks ahead of the Rio+20 summit in Brazil. Only in the areas of stopping ozone-layer depletion, removing lead in fuel, improving access to water and marine pollution research has there been any major advance in line with promised objectives, the report says.

Cancer fund

An attempt in California to create a cancer-research fund from the proceeds of tobacco taxes looked to have been narrowly defeated, but was still too close to call, after a state vote on the measure on 5 June. Proposition 29 would add a US$1 tax to every pack of cigarettes and generate around $735 million by 2013–14 for cancer research, smoking prevention and smoking cessation. The 'yes' side had garnered 49.2% of the vote, just 63,000 behind those rejecting the idea — but hundreds of thousands of postal votes are yet to be counted. See go.nature.com/yck8eu for more.

RESEARCH

Piezonuclear row

More than 1,000 scientists have signed an Internet appeal to Italy's research ministry requesting that the National Institute for Meteorological Research (INRiM) in Turin withdraw its intention to study 'piezonuclear reactions', nuclear fission that can supposedly be induced by mechanical stress such as compression of crystals. Alberto Carpinteri, a prominent proponent of

the controversial idea, was appointed president of INRiM last August, and introduced the theme into the institute's 2012–14 research plan, which was published on 29 February. On 6 June, parliamentarians also asked the research minister about INRiM's use of public funds.

Thawed brains

About 150 brains donated for research — including more than 50 from people with autism — may no longer be fit for study after a freezer malfunction at the Harvard Brain Tissue Resource Center at McLean Hospital in Belmont, Massachusetts. After the refrigeration system and two independent alarms all failed, an investigator discovered the calamity on 31 May. On 11 June, staff at Autism Speaks, a non-profit organization based in New York City that leads a programme to distribute brain samples to autism researchers, alerted the families of donors to the situation. An investigation into the freezer breakdown is to be completed by the end of June. See go.nature.com/fu14vi for more.

PEOPLE

Sci-fi writer dies

S. CASTILLO/AP

Ray Bradbury (**pictured**), the acclaimed author of the science-fiction classics *The Martian Chronicles* and *Fahrenheit 451*, among other works, died on 5 June aged 91. Bradbury's works tapped both scientific wonder and the unease about technological change, exploring how science and technology shape human identity. His stories anticipated flat-screen TVs, personal stereos and the 'butterfly effect' of chaos theory — but, in an oft-repeated remark, he said that he tried not to predict the future, but to prevent it.

Stealing secrets

A nanotechnology researcher who until recently worked at Sandia National Laboratories in Albuquerque, New Mexico, has pleaded not guilty to embezzling and passing research information to institutions in China. Jianyu Huang was charged on 5 June with five counts of federal-programme fraud from 2009, as well as falsely telling a counterintelligence officer that he would not take a Sandia-owned laptop on a trip to China. Sandia said that Huang had been fired in April for violating procedures, and his facility did not have access to classified national-security information.

Heartland fallout

The Pacific Institute in Oakland, California, has reinstated Peter Gleick as its president. Gleick had taken a leave of absence after his embarrassing admission that he had lied to obtain internal budget documents from the Heartland Institute, a libertarian think tank in Chicago, Illinois, that aims to combat climate science. In a 6 June statement, the Pacific Institute accepted Gleick's apology, and said that an independent investigation had confirmed his account of events. See go.nature.com/cjnqqa for more.

FUNDING

Nobel thrift

This year's Nobel prize recipients will split a pot of money 20% smaller than that shared by last year's laureates. In response to sluggish financial markets, which have eaten into the capital that supports the prize, the Nobel Committee said on 11 June that it would divide up 8 million Swedish kroner (US$1.1 million) between the winners of each prize to be announced in October; last year's prizes amounted to 10 million kroner apiece. It is the first time the prize has been cut in nominal value since 1949. See go.nature.com/1sml4p for more.

Telescope funding

The European Southern Observatory (ESO) announced on 11 June that its member states have given approval to start building the €1.1-billion (US$1.4-billion) European Extremely Large Telescope, although not everyone has committed their budget contributions. At a council meeting in Garching, Germany, six member states voted for the programme and four backed it pending confirmation of support by national ministries; four others are yet to agree to support the project, which is planned for Cerro Armazones in Chile's Atacama Desert. However, almost one-third of the telescope's funding is scheduled to come from Brazil, which has not yet ratified its ESO membership.

COMING UP

20–22 JUNE
Rio de Janeiro, Brazil, hosts this year's biggest environmental meeting, the United Nations Conference on Sustainable Development — 20 years after the same city hosted the original UN Earth Summit.
www.earthsummit2012.org

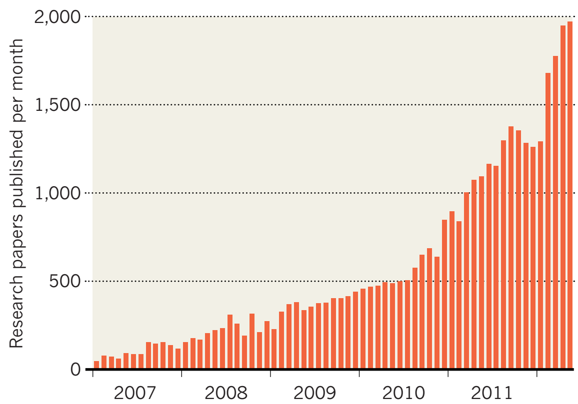
21 JUNE
The Royal Society in London launches a report on how best to open up access to modern research's vast volumes of data.
go.nature.com/1yiefm

TREND WATCH

The open-access journal *PLoS ONE* has swiftly expanded since its launch in late 2006 to become the world's largest journal; it has already published more than 9,000 articles this year (see chart). Its growth has encouraged the launch of an innovative open-access journal, *PeerJ*, which announced its business model on 12 June. But instead of charging authors per article, a one-off membership fee will give them the rights to publish peer-reviewed articles in *PeerJ* for life. See page 166 for more.

RISE OF OPEN ACCESS

The journal *PLoS ONE* expects to publish more than 2,000 research articles this month.



SOURCE: PLOS ONE

NATURE.COM
For daily news updates see:
www.nature.com/news

KENYA MILLENNIUM VILLAGES PROJECT

Health interventions may have reduced child death in African villages, but it isn't clear by how much.

GLOBAL HEALTH

Poverty project opens to scrutiny

Panel set to reshape evaluation of Millennium Villages research after partial retraction of health claims.

BY DECLAN BUTLER

The Millennium Villages Project (MVP) aims to chart a course out of poverty for the most deprived people in Africa. But after being forced to partially retract the findings of a paper that included the claim that its interventions had slashed child mortality, the project is revamping its oversight to make its science more robust and determine whether its approach works.

The MVP, a joint venture between Columbia University and the non-profit organization Millennium Promise, both in New York, is simultaneously applying a raft of aid interventions — from bed nets for malaria control to fertilizers — to help impoverished villages to meet the United Nations' eight Millennium Development Goals (MDGs), including reducing hunger, extreme poverty and the burden of major diseases, and increasing levels of primary education. The first Millennium Villages were launched in Ethiopia and Kenya in 2004 and 2005, and the project now operates in 10 African countries, reaching about 500,000 people in 80 villages over 14 sites.

➲ **NATURE.COM**
To read about how Africa is building science capacity:
nature.com/africa

In a paper published online in *The Lancet* last month[1], the project claimed a significant milestone. It reported that after three years of interventions, child mortality was decreasing three times faster in the project's villages than in the host nations in general. But the analysis was criticized for underestimating nationwide improvements in child mortality, and overestimating those in the Millennium Villages[2]. The paper's lead author, Paul Pronyk, later accepted that this claim had put the Millennium Village mortality reduction "in an excessively positive light"[3], and retracted it. Pronyk then resigned as director of Monitoring and Evaluation at Columbia University's Center for Global Health and Economic Development.

The MVP's founder, Jeffrey Sachs, head of the Earth Institute at Columbia and a co-author of the partially retracted paper, says that the MVP research teams were too autonomous, and he regrets not having brought in external advisers earlier. "I don't want such mistakes to occur again," he says. Sachs has now created a faculty committee to oversee MVP research and increase interactions with outside researchers; he will co-chair the committee, along with Cheryl Palm, a scientist at the Earth Institute's tropical agriculture programme.

An external advisory board called the International Scientific Expert Advisory Group will oversee how the MVP assesses the impact of its work. The panel is still being assembled, but will be chaired by Robert Black, head of the department of international health at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. Black says that Sachs has given him a remit "to open the study to more scrutiny, to be more transparent, to be very explicit about what are the expected outcomes of the study, and to be sure that evaluation is as well done, as rigorously, as possible". Black also wants to respond to criticisms by giving external researchers access to as much of the MVP data as he can, as well as to detailed information on the project's costs, so that outsiders can have greater confidence in the project's results and cost–benefit analyses.

It is too soon to say what sort of recommendations the external panel will make on the MVP's future research in the study villages. But given that the project is already almost seven years into its planned ten-year pilot phase, the scope for change is limited, says Black. "It is ▶

impossible to completely redesign the study, though many of us would have designed it differently," he says.

Michael Clemens, a migration and development researcher at the Center for Global Development, an independent research institution in Washington DC, is unimpressed with the creation of the advisory panel, and argues that the MVP needs rigorous and transparent evaluation from completely independent groups. Clemens, who co-authored a letter[4] to *The Lancet* that led to the partial retraction of the MVP's findings, believes that the paper still has problems — in particular, a claim that the child mortality rate in Millennium Villages fell by one-third more than that in matched comparison villages over the course of the study. "The whole study should have been retracted," he says, adding that he has submitted a new letter to *The Lancet* to this effect.

Black says that whether the MVP ultimately proves a success or failure, it will still provide useful information, so it is important to improve its evaluation mechanisms. "I don't think it should be dismissed because it has flaws," he says.

The UK government's Department for International Development last year launched its own £3.8-million (US$5.9-million) independent evaluation to accompany an £11.5-million ten-year grant to create Millennium Villages covering 30,000 people in Ghana. "A massive effort is needed to raise millions of people in Africa out of poverty. Millennium Villages represent one approach, and they could make a major contribution," the department notes. "Such an opportunity should not be missed merely for want of evidence. But nor should scaled-up resources be committed to a model without assessing whether it is indeed a cost-effective approach." ■

1. Pronyk, P. M. *et al. Lancet* **379,** 2179–2188 (2012).
2. Gilbert, N. *Nature* **485,** 158–159 (2012).
3. Pronyk, P. *Lancet* **379,** 1946 (2012).
4. Bump, J. B., Clemens, M. A., Demombynes, G. & Haddad, L. *Lancet* **379,** 1945 (2012).

RESEARCH

Journal offers flat fee for 'all you can publish'

Latest venture is part of an explosion of ideas for open-access publishing.

BY RICHARD VAN NOORDEN

Science-publishing ventures continually battle for market space, yet most operate on one of only two basic business models. Either subscribers pay for access, or authors pay for each publication — often thousands of dollars — with access being free. But in what publishing experts say is a radical experiment, an open-access venture called *PeerJ*, which formally announced its launch on 12 June, is carving out a fresh niche. It is asking its authors for only a one-off fee to secure a lifetime membership that will allow them to publish free, peer-reviewed research papers.

Relying on a custom-built, open-source platform to streamline its publication process, *PeerJ* aims to drive down the costs of research publishing, say its founders: Peter Binfield, who until recently was publisher of the world's largest journal, *PLoS ONE*, and Jason Hoyt, who previously worked at the research-paper-sharing site Mendeley. Their involvement is a major reason for the buzz around *PeerJ*. "I thought — wow — if the people I'm hearing about are working there — that's the sign of something happening. It makes it less crazy," says John Wilbanks, an advocate of open access and a senior fellow at the Ewing Marion Kauffman Foundation in Kansas City, Missouri.

PeerJ is just one of a flurry of experiments, encouraged in part by the gathering momentum of open access, that might shape the future of research publishing. "We are seeing a Cambrian explosion of experiments with new publishing models. It's going to be an interesting period for the next few years," says Binfield.

Binfield hopes *PeerJ*'s growth will resemble that of *PLoS ONE*, which went from publishing some 1,000 articles in its first full year (2007) to its current 2,000 articles a month. "*PLoS ONE* is publishing so many articles that it is stretching the boundaries of what is a journal — instead, it's becoming a large, peer-reviewed repository of research articles. We're setting ourselves up for exploring that future," says Binfield. But he adds that *PeerJ* will not need *PLoS ONE*'s volume of papers to be viable.

> *"I thought — wow — if the people I'm hearing about are working there — that's the sign of something happening."*

Whereas *PLoS ONE* charges $1,350 per paper, *PeerJ* users pay $299 for unlimited open-access publications and submissions, or a smaller fee ($199 or $99) for a limited number per year. (All authors on multi-author papers must be members, although papers with 13 or more authors need only 12 paying members.) The journal, which received undisclosed start-up support from the venture-capital fund O'Reilly AlphaTech Ventures in San Francisco, California, will be accepting articles from August.

Despite the low publication cost, *PeerJ*'s founders promise that, as with *PLoS ONE*, articles will be peer reviewed for scientific validity — but not for importance or impact. Other open-access journals have also adopted this policy, including Nature Publishing Group's *Scientific Reports*. It marks a distinction from selective open-access journals such as the forthcoming *eLife*, which plans to publish only high-impact work. To avoid running out of peer reviewers, every *PeerJ* member is required each year to review at least one paper or participate in post-publication peer review.

Untangling user fees from the publication of individual articles is a significant innovation — but other radical ideas are in the pipeline. In high-energy physics, for example, a consortium called SCOAP[3], which includes funding agencies and libraries, is planning to pay publishers for all the costs of publication, so that articles can be free to access and authors will not be charged directly. On 1 June, the SCOAP[3] initiative said that it had sent out tenders to publishers to bid for these contracts, with services expected to start in January 2014.

Other ideas under discussion include journals that charge for submissions rather than for publications; direct government funding for all publications; and research funders setting up their own publication infrastructure (much as some do with biology databases), says Cameron Neylon, recently appointed director of advocacy at the Public Library of Science in San Francisco, which publishes *PLoS ONE*.

No one knows what will work. But many say that the experiments now under way will help to reveal the true costs of sustainably publishing articles and research data. "*PeerJ* is part of the assertion that this can be done cheaper — and for that alone it will be interesting to watch," says Neylon. ■

The rise of the right: Hungary's Jobbik party rallies supporters in Budapest.

T. KOVACS, MTI/AP

GENETICS

Genome test slammed for assessing 'racial purity'

Hungarian far-right politician certified as 'free of Jewish and Roma' genes.

BY ALISON ABBOTT

Officials in Hungary united this week to condemn ongoing ethnic violence and anti-Semitic attacks, including an assault on the former Chief Rabbi on 5 June. But a cause for further soul-searching has emerged: a scientific scandal recalling discredited notions of racial purity.

Hungary's Medical Research Council (ETT), which advises the government on health policy, has asked public prosecutors to investigate a genetic-diagnostic company that certified that a member of parliament did not have Roma or Jewish heritage.

The MP in question is a member of the far-right Jobbik party, which won 17% of the votes in the general election of April 2010. He apparently requested the certificate from the firm Nagy Gén Diagnostic and Research, which rents office space at the prestigious Eötvös Loránd University in Budapest. The company produced the document in September 2010, a few weeks before local elections.

The certificate — with the MP's name blacked out — emerged on the web last month and was seized on by the Hungarian media. One of Nagy Gén's financial partners, Tibor Benedek — a three-time Olympic water-polo gold medallist and a member of a prominent Jewish family — immediately pulled out of the company.

The ETT's secretary, József Mandl, chair of medical chemistry at the Semmelweis University in Budapest, says that the certificate is "professionally wrong, ethically unacceptable — and illegal". The council discussed the issue on 7 June and concluded that the genetic test violates the 2008 Law on Genetics, which allows such testing only for health purposes.

"The council's stand is important," says Lydia Gall, an Eastern Europe and Balkans researcher at civil-rights group Human Rights Watch, who is based in Amsterdam. In Hungary, "there have been many violent crimes against Roma and acts of anti-Semitism in the past few years", she says. Politicians who try to use genetic tests to prove they are 'pure' Hungarian fan the flames of racial hatred, she adds.

Nagy Gén scanned 18 positions in the MP's genome for variants that it says are characteristic of Roma and Jewish ethnic groups; its report concludes that Roma and Jewish ancestry can be ruled out. The certificate adds: "For an interpretation of the test result and for genetic consultation relating to the family-tree research, please contact us as soon as convenient."

Nagy Gén did not respond to e-mail and telephone requests from *Nature* for comment. But a statement on its website claims that newspapers had reported the story "incompletely" and points to the certificate's recommendation for further consultation. It argues that the company "rejects all forms of discrimination, so it has no right to judge the purpose for which an individual will use his or her test result, and so for ethical reasons it could not have refused to carry out the test".

The certificate first appeared on a right-wing website, which described the intention behind the gene test as "noble", although it questioned the science. After the news blog Petőfi utca republished the certificate on 14 May, the Hungarian Society of Human Genetics issued a statement condemning the test. István Raskó, director of the Institute of Genetics of the Hungarian Academy of Sciences in Szeged, and the society's vice-president, says that it is impossible to deduce origins from genetic variations at a few places in the genome. "This test is complete nonsense and the affair is very harmful to the profession of clinical genetics," he says.

Nagy Gén's rental contract with Eötvös Loránd University ended this month, says György Fábri, a university spokesman. "The university is not commenting publicly on the affair because it is not our business — our researchers had no contact with the company." In a written statement he added that the university "fully rejects" the abuse of scientific results to promote discrimination or hatred. ■

The United Kingdom Infra-Red Telescope is setting records for scientific publications.

ASTRONOMY

Stellar UK scope faces closure

Specialize or die is the mantra for medium-sized instruments.

BY ERIC HAND

Perched near the 4,200-metre summit of Mauna Kea in Hawaii, the United Kingdom Infra-Red Telescope (UKIRT) enjoys one of the finest sites in all of astronomy. It was among the world's largest telescopes when it opened more than 30 years ago, and its enviable record of publications shows no signs of slowing. Yet on 30 May, the UK Science and Technology Facilities Council (STFC) announced that it will close the 3.8-metre telescope in 2013, unless a buyer can be found.

UKIRT's closure could be the first of many, say concerned astronomers. Just as 1- and 2-metre telescopes began to be mothballed in the 1990s to free up funding for a new class of 8–10-metre instruments, more than a dozen 4-metre telescopes around the world could face a similar fate as astronomers prepare for the arrival of 30-metre behemoths (see *Nature* **479,** 18–19; 2011).

"Is this a sign of things to come? It may be," says Gary Davis, director of the Joint Astronomy Centre in Hawaii, which operates UKIRT. The cash-strapped STFC decided to sacrifice UKIRT to preserve its membership commitment to the European Southern Observatory (ESO), which is planning a 40-metre giant. "Once again, the UK is leading the way," Davis adds archly.

But survival strategies are coming to light. Chief among them is that, rather than using 4-metre telescopes as general-purpose observatories, with time being divided among hundreds of astronomers, they should be dedicated to specific problems requiring large collaborations, long campaigns and custom-built instruments. "We've become much more like particle physics, oriented towards doing specific things in big teams," says Janet Drew, an astronomer at the University of Hertfordshire in Hatfield, UK, who chaired a committee planning the future of Europe's 2–4-metre telescopes.

In the United States, for example, the Department of Energy is hoping to fund the development of instruments that will allow two ageing, 4-metre-class telescopes belonging to the National Optical Astronomy Observatory (NOAO) to engage in a quest to understand dark energy, the mysterious force that is speeding up the expansion of the Universe.

➲ NATURE.COM
Read more about 30-metre telescopes at:
go.nature.com/tlvntb

One collaboration, the Dark Energy Survey, will install a 570-megapixel camera on the 4-metre Blanco telescope in Chile. By taking pictures of hundreds of millions of galaxies, it could reveal tiny distortions in their shape that would betray the effect of dark energy. Another group, called BigBOSS, plans to investigate dark energy by looking for patterns in the distribution of 20 million galaxies using an instrument installed at the 4-metre Mayall telescope in Arizona (see *Nature* **481,** 10–11; 2012). "The highest-impact science right now are these things where you want to put in more than 100 nights a year," says NOAO director David Silva.

In Europe, one productive niche has been the search for planets outside the Solar System. An instrument at the ESO's 3.6-metre telescope at La Silla in Chile has already discovered more exoplanets than any other ground-based search. And in April, the University of Geneva in Switzerland installed a near-clone of that instrument in the Northern Hemisphere, at Italy's 3.6-metre National Galileo Telescope in the Canary Islands.

A third emerging use for these telescopes is to obtain spectra of millions of stars in the Milky Way, to try to understand their chemistry and their wanderings through space. This would allow astronomers to reconstruct a detailed history of the Galaxy's formation. A spectrograph specialized for this purpose is already being built for the 3.9-metre Anglo-Australian Telescope in Australia, and the ESO is considering one for its 3.6-metre New Technology Telescope in Chile.

But Drew says that it is sometimes difficult to convince disparate user groups to coalesce around a coordinated scientific campaign. And, especially in Europe, multinational ownership makes it difficult for telescope operators to change course. "Structures in Europe are oddly complex," she says. "They get in the way a bit."

Moreover, UKIRT itself is proof that a dedicated speciality campaign does not guarantee survival. In 2005, for example, its operators invested in a wide-field camera that allowed it to survey galaxies across the sky. Patrick Roche, an astronomer at the University of Oxford, UK, and chair of UKIRT's oversight board, says that in the first four months of this year, more than 70 papers have been published containing UKIRT data. That puts UKIRT on course to being the most scientifically productive telescope in the world — outstripping any of the 8-metre-class instruments.

The telescope's board, which has protested against the closure decision, claimed that the facility could be kept open a year longer for a mere £100,000 (US$155,000). And Davis thinks that he might be able to find an institution willing to take on the costs, although he hasn't quite settled on the best way to advertise a 34-year-old telescope with a fine 3.8-metre mirror.

"eBay?" he wonders. "I don't know." ■

FUNDING

Grant review opens up Texas-sized rift

Big money and big ambitions roil state's research efforts.

BY MEREDITH WADMAN

When Ronald DePinho took the helm at the MD Anderson Cancer Center in Houston, Texas, last September, he wasn't shy about his ambitions. He is on a "Moon shot" quest, he said, with the aim of making huge gains against five cancers in the next decade, partly by attacking a void in cancer drug development. Wasting no time, DePinho launched a drug-discovery centre, the Institute for Applied Cancer Science (IACS), where his wife, Lynda Chin, was appointed chief scientist. The effort gained momentum in March, when an IACS research team led by Chin won most of a US$20-million commercialization grant from the taxpayer-funded Cancer Prevention and Research Institute of Texas (CPRIT), based in Austin.

But within weeks the funding coup had turned into a media firestorm — one that has led to harsh criticism of the CPRIT's decision-making process and exposed tensions within the community the institute serves. The affair led the CPRIT to announce on 6 June that it plans to re-review the award it had made to Chin, this time on its scientific merits. The original review had considered only its commerical potential. And this week the embattled institute has advertised for its first compliance officer — whose job it will be to ensure that the institute's internal policies and procedures for grant-making are followed.

"Hindsight is twenty–twenty," says the CPRIT's executive director William Gimson. "We want to hold ourselves to a higher standard. So we are willing to come back and say: 'If these questions have been raised, let's have a scientific and a commercial review'."

The controversy became public with the revelation on 8 May that the CPRIT's chief scientific officer, Nobel laureate Alfred Gilman, was resigning. Gilman, it emerged, was incensed by the speedy awarding of the IACS grant without a scientific review and by the simultaneous shelving of seven research awards worth $39 million that the institute's scientific reviewers had recommended for funding. If the situation was not fixed, and quickly, Gilman warned in an e-mail to top managers obtained under Texas public-records law, "cancer patients will lose; the citizens of Texas will be deceived; the integrity of science in Texas will be soiled".

The state has been the envy of cancer scientists since 2007, when its citizens voted overwhelmingly to establish the CPRIT and authorized $3 billion in bond issues to support its efforts. The money — $300 million a year for ten years — began flowing in 2010. Since then, CPRIT has funded 387 grants worth a total of $671 million — most of them to academic research institutions (see 'Mapping the millions'). Roughly 17% of the CPRIT's funds have gone to commercialization, helping Texas companies to develop cancer diagnostics and treatments. Before the IACS award, all grant applications were subject to both commercial and scientific review.

F. CARTER SMITH

"The reality is: we applied, our proposal was reviewed and it got funded."
Ronald DePinho

Then, in September, as DePinho and Chin arrived in Houston from the Dana-Farber Cancer Institute in Boston, Massachusetts, the CPRIT unveiled a different species of commercial grant. Called an 'incubator' grant, its aim was to provide expert technical and business assistance to Texas-based translational research projects and start-ups. The IACS team seized on the opportunity with a six-and-a-half-page proposal and submitted its application on 11 March. By 29 March, an annually renewable grant worth $18 million per year to the IACS and $2 million to Rice University in Houston had been approved by the CPRIT's Commercialization Review Council and signed off by the institute's governing Oversight Committee. The grant named Chin as principal investigator.

On 24 April, after seeing a copy of the grant proposal, Gilman wrote to the CPRIT's Scientific Review Council, which is chaired by Nobel laureate Phillip Sharp at the Massachusetts Institute of Technology in Cambridge, complaining about the award. In Gilman's view, the proposal was a "back door" submission, amounting to no more than "a vague organizational plan" "You would be shocked to see it," he wrote.

After Gilman resigned, Gimson received a blistering missive from the eight members of the scientific council. The institute's failure ▶

to include them in reviewing the IACS award "calls into question our roles and the integrity of the review program in general", they wrote. "This by-pass is inherently unfair to every scientist in Texas who participates in the CPRIT program … These scientists have played by the rules."

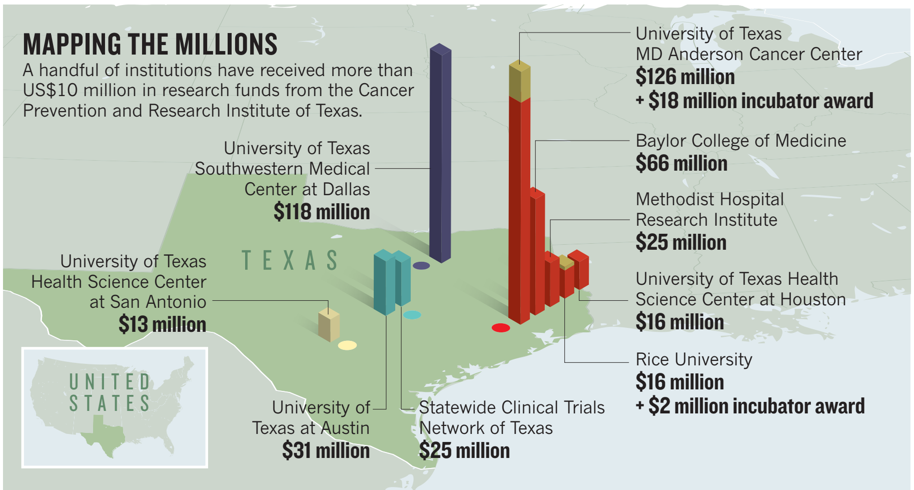
Both Chin and DePinho have repeatedly insisted that they, too, played by the rules. DePinho, who asked the CPRIT if it wanted to re-review the grant after the controversy erupted, says that he is certain that the award will stand on its own merits and notes that CPRIT officials encouraged the MD Anderson team to try for the award. "The reality is: we applied, our proposal was reviewed and it got funded," he says, adding that the team's submission "was done in a way that was totally consistent with the CPRIT's guidelines". MD Anderson provost Raymond DuBois and Kenneth Shine, the University of Texas System vice-chancellor for health affairs, to whom DePinho reports, say that they agree.

Yet CPRIT correspondence suggests that some in authority at the institute worked behind the scenes to make sure that the MD Anderson team qualified for the grant. For instance, the minutes of an internal telephone conference sent on 4 January to Robert Ulrich, chairman of the commercialization council, and Jerry Cobbs, CPRIT's chief commercialization officer, includes an action item for Cobbs to "change some wording" to the incubator grant request for applications, "to broaden the scope to include incubators that would serve earlier-stage programs, such as Lynda Chin's endeavor."

HKSTRATEGIES

"The integrity of science in Texas will be soiled."
Alfred Gilman

The compliance office of the University of Texas System, of which MD Anderson is a part, is investigating why the incubator grant application was submitted directly from Chin's team without first being reviewed by the MD Anderson provost. The provost normally reviews all grant applications going out of the institution, looking, in part, for financial conflicts, or appearances of them. "I do have some concern about the precise pathway by which the grant was submitted," says Shine. DePinho acknowledges that the application should not have been submitted directly to the CPRIT.



SOURCE: CPRIT

The *Houston Chronicle* has also raised questions about potential conflicts of interest among some of the application's commercial reviewers. The CPRIT repudiates this, noting that those with connections to DePinho or Rice University recused themselves from voting on the relevant parts of the award.

Nonetheless, appearances can be problematic, says Paul Root Wolpe, director of the Center for Ethics at Emory University in Atlanta, Georgia. "An absolutely unassailable process is desirable," he opines. "Greater pains should have been taken to avoid any appearance of favouritism."

DePinho did not improve perceptions when, on 18 May, he was a guest on a national cable television show that offers investment tips. When asked which cancer drug companies would make good buys, he touted AVEO Oncology of Cambridge, mentioning that he had helped to found the company but not that he owns 542,000 shares in it. DePinho later apologized for his comments in an article in the weekly newsletter *The Cancer Letter*. Shine says that a conflict-monitoring committee for the University of Texas System is examining the incident. "We will be taking a look at that episode and what it has to teach us."

The controversy has also exposed some of the competing agendas within CPRIT. E-mails show that DePinho's ambitious style has irked some scientists who are connected with the CPRIT award process and that some of the Oversight Committee members were unhappy that so many grants were flowing to the University of Texas Southwestern Medical Center at Dallas, where Gilman is based. Gilman wrote to CPRIT executives that the most meritorious proposals were being funded, irrespective of geography, adding: "We have assembled the best from outside the state to evaluate our programs. If we do not accept their advice we are arrogant and foolish."

Gilman's e-mails suggest that he was increasingly at odds with the Oversight Committee, which consists mainly of political appointees. In his letter of resignation, he wrote that he wanted stay until October, which would, among other things, allow him to monitor a crucial 26 July meeting of the committee, when the seven grants sidelined in March, including five to the University of Texas Southwestern Medical Center worth a total of nearly $26 million, will again come up for approval. Gimson will also seek the committee's approval to review the MD Anderson grant and others like it using a merged committee — half commercialization experts and half scientific reviewers — and to require from MD Anderson a level of scientific detail not present in its original proposal.

Noting that this is the first of hundreds of CPRIT grants to have raised questions, Gimson says: "From my perspective we have had a stellar record. My insistence will be that we continue to have that stellar record." ■

MORE ONLINE

TOP STORY

EDELMANN/SPL

Fetal genome deduced from parental DNA go.nature.com/zsylkq

NEWS

- NASA scientists raise awareness of budget cuts with cupcakes go.nature.com/b95na3
- Vaccines used to control cholera outbreak in Africa go.nature.com/eitejb
- Vitamin D fails diabetes test go.nature.com/onhrfn

VIDEO

M. MEUR/STOCKTREK IMAGES/CORBIS

Giant reef fish defend territory with headbutts go.nature.com/baelyu

PHYSICS

Tabletop X-rays light up

Compact device promises to open window on chemical reactions in the lab.

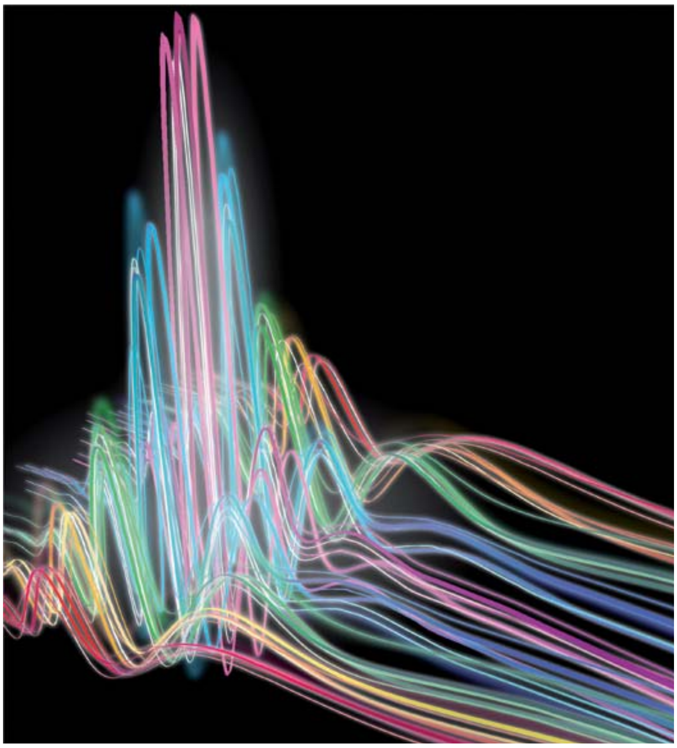
BY KATHERINE BOURZAC

The pressurized, cylindrical chamber fits in the palm of Margaret Murnane's hand. Yet out of one end of the device comes an X-ray beam that packs almost as much punch as the light generated by massive particle accelerators.

Murnane and Henry Kapteyn, both physicists at JILA in Boulder, Colorado, a joint institute of the University of Colorado and the US National Institute of Standards and Technology, have reported the first tabletop source of ultra-short, laser-like pulses of low energy, or 'soft', X-rays. The light, capable of probing the structure and dynamics of molecules, was previously available only at large, billion-dollar national facilities such as synchrotrons or free-electron lasers, where competition for use of the equipment is fierce. But the report by Murnane, Kapteyn and their colleagues, published in the 8 June issue of *Science* (T. Popmintchev *et al. Science* **336,** 1287–1291; 2012), suggests that the devices might soon lie within the grasp of a university laboratory budget. "For us, it's incredible that we can do this at all in a tabletop system," says Murnane. "Three years ago, people would have said 'only large facilities can do that'."

Murnane and Kapteyn, a husband-and-wife team who also head the Boulder-based company KMLabs, already sell a similar tabletop source of extreme-ultraviolet light. Murnane thinks that a future soft X-ray source should cost about US$1 million, and hopes that its relatively low cost and small size will open up X-ray studies for materials scientists, biologists and others. The beams generated by the device could, for instance, help materials scientists to make better solar materials by tracking the paths of electrons through solar cells, and might allow chemists to trace the ultrafast dynamics of photosynthesis and catalysis. "This is something people have been waiting for for a very long time," says Oleg Shpyrko, a physicist at the University of California, San Diego. Shpyrko often waits months to get his experiments accepted at the Advanced Photon Source, a synchrotron at Argonne National Laboratory in Illinois, and must then fly his students halfway across the country to do the work.

The tabletop sources rely on a technique called high-harmonic generation, in which laser light is passed through a medium that converts it to light of shorter wavelengths and higher frequencies. Shine a ruby laser into a quartz crystal, for example, and a beam of ultraviolet light comes out — albeit dimmer, but still focused like a laser beam.



Chemical bonds can be probed with fast, laser-like X-ray pulses (illustrated).

T. POPMINTCHEV & B. BAXLEY/JILA/UNIV. COLORADO

Murnane and Kapteyn have pushed high-harmonic generation to its limits, with a system that uses an infrared laser as the source and pressurized helium gas as the medium. The laser creates a strong electric field, which draws electrons away from the helium atoms, allowing the electrons to absorb energy from the electric field. When they slam back in to the helium atoms, they release that absorbed energy as shorter-wavelength photons — but only about one photon comes out for every 5,000 infrared photons that are put in.

The result is light at wavelengths almost as short as those delivered by synchrotrons. By increasing the pressure in the gas — something theorists thought might defocus the light beam — Murnane and Kapteyn can produce light with a wavelength of 0.8 nanometres. Around these wavelengths, many of the chemical elements used in magnets and superconductors absorb a characteristic band of light. Kapteyn says that this could be used by chemists to discern the spin states of nickel atoms that make up the bits of information in magnetic computer hard drives, for example.

The tabletop systems have already surpassed the larger light-source facilities in terms of laser pulse speed. Murnane and Kapteyn's device produces very fast pulses, as short as 2.5 attoseconds (10^{-18} seconds) — faster than the picosecond (10^{-12}) pulses of synchrotrons and the femtosecond (10^{-15}) pulses of free-electron lasers. That timescale is even quicker than the making and breaking of chemical bonds. "At this timescale, we can start using these light sources to address questions where we have no idea what the answers will be," says Ferenc Krausz, a physicist at the Max Planck Institute for Quantum Optics in Garching, Germany, and co-founder of Femtolasers, a company that sells ultrafast light sources.

Light produced by these sources is much less intense than that at the big national facilities, however. That's one reason why physicist Emma Springate, who runs Artemis, part of the Central Laser Facility at the Science and Technology Facilities Council's Rutherford Appleton Laboratory near Didcot, UK, wants access to both technologies. Artemis already has one of KMLabs' ultrafast extreme-ultraviolet sources, which it combines with access to a synchrotron source. "The synchrotron gives you a static, really clear high-resolution picture, and the ultrafast source gives you a slightly fuzzy movie," Springate says.

Tabletop X-ray sources could still be several years away, Murnane says, but she hopes that they will one day be as common in labs as electron microscopes are. Shpyrko, for one, is looking forward to that day. Waiting to get an experiment accepted at a national facility, he explains, can feel like waiting for a space shuttle to launch. "If you have one of these in your own lab, you can dream up an experiment and try it tomorrow," he says. ■

NEWS FEATURE

Samples crushed between two diamonds can be probed with lasers.

HARD PRESSED

TWO PHYSICISTS SAY THEY HAVE FORCED HYDROGEN TO BECOME AN EXOTIC METAL THOUGHT TO EXIST ONLY IN THE HEARTS OF GIANT PLANETS. NOW THEY MUST FACE THEIR CRITICS.

BY IVAN AMATO

Mikhail Eremets and Ivan Troyan work only with the very best, gem-quality diamonds. Nothing else can handle the stresses involved — not when the two physicists make a habit of forcing the diamonds' tips together until the pressure between them reaches levels normally found at the centre of the Earth.

Eremets and Troyan, both at the Max Planck Institute for Chemistry in Mainz, Germany, are not alone in this strange pursuit. Their apparatus, known as a diamond anvil cell, is a standard fixture in high-pressure research labs. But they have provoked widespread agitation in their research community by claiming to have crushed hydrogen in their cells until it gave up being a diatomic gas, and instead became a shiny, presumably monatomic, solid that conducted electricity like a metal.

The arguments over this alleged discovery — one of the most sought-after results in high-pressure research — have been raging ever since Eremets and Troyan published their results last November[1]. "We could just read their paper and say it was wrong," declares condensed-matter physicist William Nellis, an associate of Harvard University's physics department in Cambridge, Massachusetts, and one of the team's most vocal challengers[2].

But others, such as Raymond Jeanloz, a planetary scientist and high-pressure-materials researcher at the University of California, Berkeley, are keeping an open mind. Whether Eremets and Troyan are eventually proved right or wrong about metallic hydrogen, Jeanloz says, "what I feel is beautiful about their work is that they did a bazillion different experiments at these extreme conditions. They lit a fire under a couple of research groups."

BEN SMITH

On 26 June, the main figures in this controvery will gather in Biddeford, Maine, during a Gordon Research Conference on high-pressure science. The result could be a meeting of minds — or a display of fireworks. "This is a very intense field," says Jeanloz. High-pressure scientists have been trying — and failing — to make metallic hydrogen ever since theorists first predicted its existence in 1935 (ref. 3). Anyone claiming success can expect an all-out critique from rival groups — especially considering what is at stake.

Not only would making metallic hydrogen in the laboratory allow researchers to do planetary science at the bench — gas-giant planets such as Jupiter, or the even larger ones being discovered around distant stars, are thought to have huge amounts of the stuff in their interiors — but it could point the way towards an entirely new world of high-pressure phenomena.

"Hydrogen is the simplest atom, the simplest molecule and perhaps the most complicated elemental solid," says Arthur Ruoff, a high-pressure physicist at Cornell University in Ithaca, New York. In 1968, Cornell physicist Neil Ashcroft predicted that solid metallic hydrogen might be a superconductor[4]. In 2004, calculations[5] by Ashcroft and others suggested that, under certain combinations of pressure and temperature, hydrogen atoms would rearrange themselves into a new kind of quantum liquid with attributes of both superconductors, which conduct electricity without resistance, and superfluids, which flow without resistance.

BRIGHT FUTURE

Such exotic behaviours become particularly interesting if some of these phases of hydrogen turn out to be metastable. This would mean that the phases could retain their high-pressure forms for an indefinite period once external forces are removed, much as diamonds formed by high temperatures and pressures deep inside Earth remain diamonds even after they reach the surface, instead of immediately reverting to carbon's more stable form, graphite. Nellis and others have imagined a host of applications for metastable metallic hydrogen, ranging from super-lightweight structural materials that would allow entire cities floating on the sea to be built, to rocket fuel that packs nearly four times as much propellant power per kilogram as the liquid hydrogen used in the most powerful rockets today[6].

First, however, comes the reality check. Diamond anvil cells can use only vanishingly small sample sizes. (The volume of Eremets and Troyan's hydrogen samples was about 160 cubic micrometres — somewhat smaller than an average human cell.) High-pressure experiments are fraught with the potential for error. And even the most experienced researchers run the risk of fooling themselves. The question is whether Eremets and Troyan are among them.

In their experiments[1], the Mainz physicists started with pairs of brilliant-cut diamonds. They trimmed each of the diamonds' points into a flat surface, or culet, 20–30 micrometres across, then aligned the diamonds with those truncated tips almost touching, on either side of a piece of metal foil pierced with a culet-sized hole that would enclose a minuscule experimental chamber.

THROUGH A GEM DARKLY

For each experiment, Eremets and Troyan loaded hydrogen gas into the hole, and started tightening a set of screws that forced the diamonds closer together. As the culets bit into the foil at the rim of the hole, the metal deformed around them to form a seal trapping the hydrogen. And as the force exerted by the screws was focused down onto the culets, the pressure in the chamber began to skyrocket.

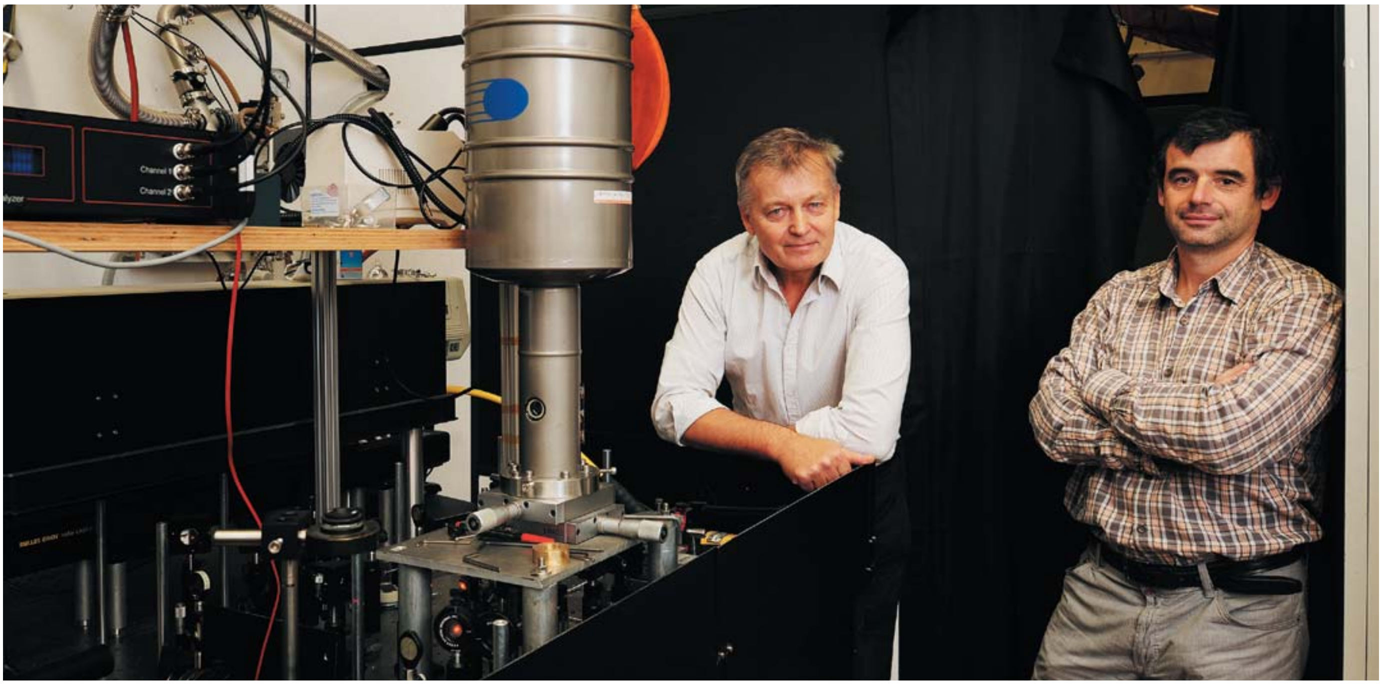
This is where high-pressure experiments often go awry. Although the anvil's diamond jaws are made of one of the hardest materials known, superpressured hydrogen routinely infiltrates them, making them brittle and causing the gems to crack like cheap plastic cups. This also spoils the gemstones' transparency, stopping the physicists from seeing what is happening in the chamber. Eremets and Troyan addressed that problem by applying semi-transparent coatings to protect their diamonds from the infiltration of hydrogen. But even so, they went through about 100 pairs of diamonds as they worked out the bugs of their apparatus and ran trials and controls.

When the pressure inside their anvils reached 200 billion pascals (200 GPa), or about 2 million times Earth's atmospheric pressure at sea level, the two Mainz physicists saw spectroscopic signs that the diatomic hydrogen molecules had begun to interact in a way that signalled that they were becoming a solid.

At 220 GPa, the sample chamber became dark, another sign that hydrogen had apparently assumed a condensed phase. Eremets and Troyan found that laser pulses fired through the transparent diamond triggered small flows of electrons that they could detect by means of copper and gold electrodes deposited on the culet surfaces — behaviour characteristic of semiconducting materials, in which electrons need a small energy kick to flick them out of orbit around atoms and into a conductive flow.

At 240 GPa, Eremets and Troyan recorded small currents even without the laser shots, an indication that room-temperature thermal vibrations alone could nudge the electrons into a conductive mood.

And at about 270 GPa, the researchers saw their samples' electrical resistance suddenly drop by several orders of magnitude, just as their spectroscopy was showing that the hydrogen's molecular vibrations were slowing



Mikhail Eremets (left) and Ivan Troyan stand by the apparatus they used to create what they believe to be metallic hydrogen.

CARSTEN COSTARD

down and morphing in ways suggestive of a phase transition — and the samples were becoming shiny like a metal.

"It is hard to sit on data like these," says Eremets. He and Troyan are convinced that their findings are at least consistent with having made metallic hydrogen.

But others are not so sure. Sceptics note that pitfalls are everywhere in experiments such as these. If an experimenter peers through the diamonds and sees the sample begin to darken, for example, that may mean that it is becoming a solid — or that the hydrogen is reacting with the foil or other impurities to form metal hydrides. If odd electrical signals start coming in from the culet electrodes, they may indicate the formation of a new phase of hydrogen — or that the diamond jaws are deforming, causing the electrodes to short out or emit spurious signals. Eremets says that he and Troyan are well aware of such pitfalls, and did their best to avoid them. Nonetheless, says Ruoff, "I don't know anybody who thinks their claim is valid".

One fact that troubles Ruoff and other critics is that the resistance rises as the sample's temperature drops. This is contrary to normal metallic behaviour, in which the resistance goes down as the temperature does. But when reviewers of the paper raised that issue, says Eremets, he successfully argued that the upward trend is consistent with that seen in a disordered metal, in which the atomic structure is more like that of a glass than a crystal.

Russell Hemley, a high-pressure-materials researcher at the Carnegie Institution of Washington in Washington DC, is sceptical of Eremets and Troyan's claims for another reason: if the hydrogen really becomes metallic, then the shininess that the Mainz team reports at optical wavelengths ought to be accompanied by equally strong reflectivity at infrared wavelengths. Eremets and Troyan were not able to get good infrared data on their samples. But Hemley says that in his own experiments with hydrogen at similar pressures, "we see it transmitting" in the infrared. Moreover, in a paper published in April[7], Hemley and his colleagues report signs that hydrogen squeezed up to 360 GPa holds on to its diatomic character and fails to morph into a monatomic metal. Adding to the complexity of the situation, some theorists suggested[8] that, if squeezed enough, even intact hydrogen molecules might develop a new and crowded bonding pattern that could have metallic properties.

"OUR MEASUREMENTS ARE NOT PERFECT, BUT METALLIC HYDROGEN REMAINS A VIABLE INTERPRETATION."

Eremets and Troyan are standing their ground. "Our measurements are not perfect," Eremets concedes, but he insists that "metallic hydrogen remains a viable interpretation". He points to a paper published in March[9] by Eugene Gregoryanz, a physicist at the University of Edinburgh, UK, and his colleagues. "They observe a new phase, the same Raman spectra and a darkening of the sample" at around 220 GPa, says Eremets. Clearly, he argues, something is happening at that point — although Gregoryanz and his colleagues attributed the changes to formation, not of metallic hydrogen, but of a previously unseen, graphene-like phase of molecular hydrogen.

In an effort to clarify the situation, Eremets and Troyan travelled to Villigen in Switzerland in late April, to gather infrared spectra using the synchrotron light source at the Paul Scherrer Institute there. The data could reveal vibrations in the lattice structures of high-pressure hydrogen samples that would help theoreticians to identify specific structures of any solid phases that might be present, Eremets says. Such structural information would not, by itself, prove that the hydrogen had become metallic. But Eremets and Troyan also plan to attempt more thorough and extensive electrical measurements, which could prove conclusive.

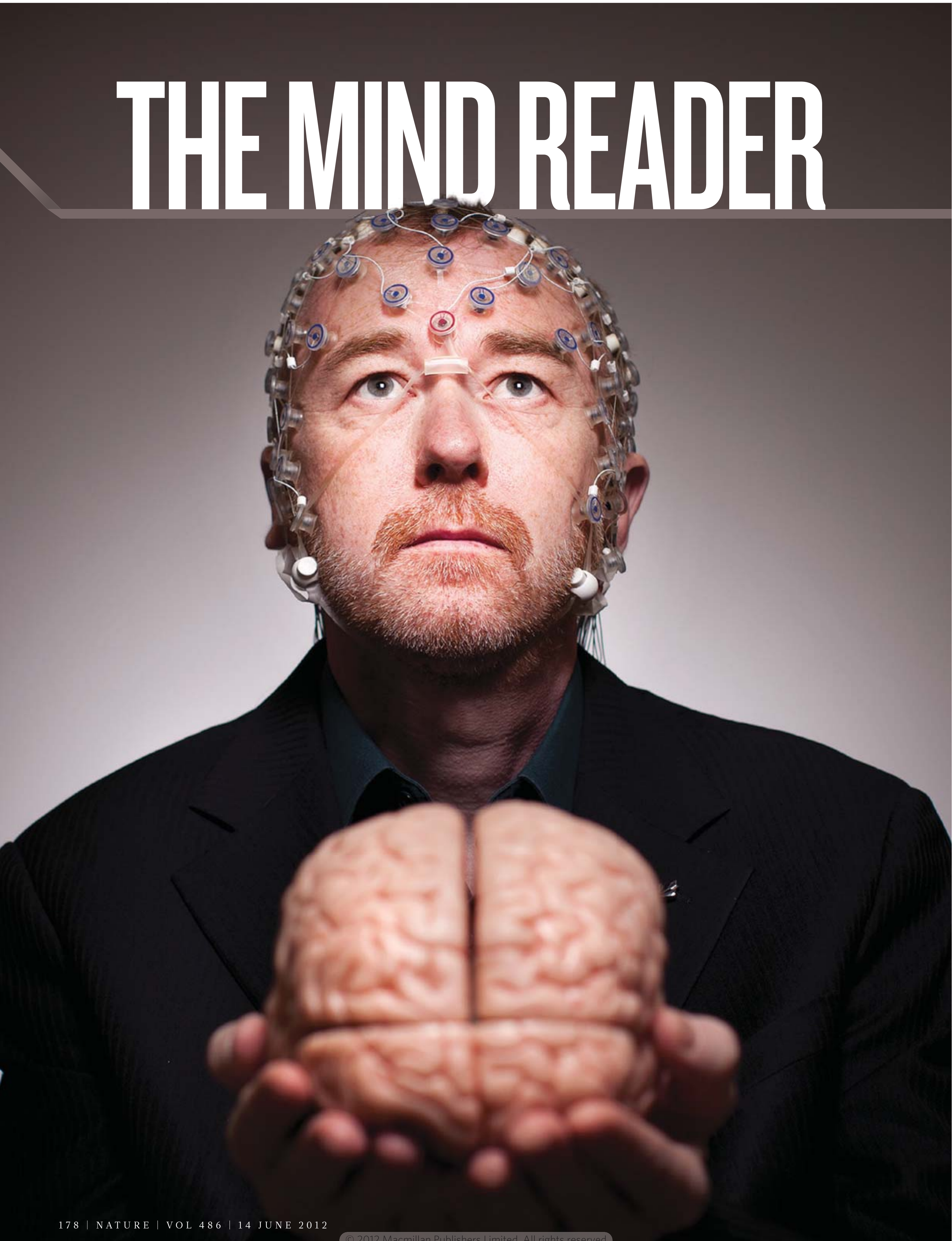
Looking back, Eremets admits that he and Troyan probably should have been more circumspect in their paper — conveying the message that they might have made metallic hydrogen, rather than claiming more strongly that they had done so, as they did in their abstract and their concluding paragraphs.

Meanwhile, the race for metallic hydrogen continues, as other high-pressure research groups pursue their own experiments. Jeanloz points out that many members of this community, himself included, have mingled over the years as collaborators, postdocs, supervisors and mentors, only to go on to become each other's staunchest critics and rivals, unlikely to let the slightest chink in an argument go unchallenged. What emerges from this dynamic, say Jeanloz and others, is a creative tension that could eventually force those who claim to have produced metallic hydrogen to do enough experiments, with enough controls, to compile enough lines of evidence to convince even the pickiest of critics. Until then, the trophy seems to be up for grabs. ■

Ivan Amato *is a freelance writer based in Silver Spring, Maryland.*

1. Eremets, M. I. & Troyan, I. A. *Nature Mater.* **10,** 927–931 (2011).
2. Nellis, W. J., Ruoff, A. L. & Silvera, I. F. Preprint at http://arxiv.org/abs/1201.0407 (2012).
3. Wigner, E. & Huntington, H. B. *J. Chem. Phys.* **3,** 764–770 (1935).
4. Ashcroft, N. W. *Phys. Rev. Lett.* **21,** 1748–1749 (1968).
5. Babaev, E., Sudbø, A. & Ashcroft, N. W. *Nature* **431,** 666-668 (2004).
6. Silvera, I. F. & Cole, J. W. *J. Phys. Conf. Ser.* **215,** 012194 (2010).
7. Zha, C.-S., Liu, Z. & Hemley, R. J. *Phys. Rev. Lett.* **108,** 146402 (2012).
8. Labet, V., Gonzalez-Morelos, P., Hoffmann, R. & Ashcroft, N. W. *J. Chem. Phys.* **136,** 074501 (2012).
9. Howie, R. T. *et al. Phys. Rev. Lett.* **108,** 125501 (2012).

THE MIND READER



Adrian Owen has found a way to use brain scans to communicate with people previously written off as unreachable. Now, he is fighting to take his methods to the clinic.

BY DAVID CYRANOSKI

Adrian Owen still gets animated when he talks about patient 23. The patient was only 24 years old when his life was devastated by a car accident. Alive but unresponsive, he had been languishing in what neurologists refer to as a vegetative state for five years, when Owen, a neuroscientist then at the University of Cambridge, UK, and his colleagues at the University of Liège in Belgium, put him into a functional magnetic resonance imaging (fMRI) machine and started asking him questions.

Incredibly, he provided answers. A change in blood flow to certain parts of the man's injured brain convinced Owen that patient 23 was conscious and able to communicate. It was the first time that anyone had exchanged information with someone in a vegetative state.

Patients in these states have emerged from a coma and seem awake. Some parts of their brains function, and they may be able to grind their teeth, grimace or make random eye movements. They also have sleep–wake cycles. But they show no awareness of their surroundings, and doctors have assumed that the parts of the brain needed for cognition, perception, memory and intention are fundamentally damaged. They are usually written off as lost.

Owen's discovery[1], reported in 2010, caused a media furore. Medical ethicist Joseph Fins and neurologist Nicholas Schiff, both at Weill Cornell Medical College in New York, called it a "potential game changer for clinical practice"[2]. The University of Western Ontario in London, Canada, soon lured Owen away from Cambridge with Can$20 million (US$19.5 million) in funding to make the techniques more reliable, cheaper, more accurate and more portable — all of which Owen considers essential if he is to help some of the hundreds of thousands of people worldwide in vegetative states. "It's hard to open up a channel of communication with a patient and then not be able to follow up immediately with a tool for them and their families to be able to do this routinely," he says.

Many researchers disagree with Owen's contention that these individuals are conscious. But Owen takes a practical approach to applying the technology, hoping that it will identify patients who might respond to rehabilitation, direct the dosing of analgesics and even explore some patients' feelings and desires. "Eventually we will be able to provide something that will be beneficial to patients and their families," he says.

Still, he shies away from asking patients the toughest question of all — whether they wish life support to be ended — saying that it is too early to think about such applications. "The consequences of asking are very complicated, and we need to be absolutely sure that we know what to do with the answers before we go down this road," he warns.

LOST AND FOUND

With short, reddish hair and beard, Owen is a polished speaker who is not afraid of publicity. His home page is a billboard of links to his television and radio appearances. He lectures to scientific and lay audiences with confidence and a touch of defensiveness.

Owen traces the roots of his experiments to the late 1990s, when he was asked to write a review of clinical applications for technologies such as fMRI. He says that he had a "weird crisis of confidence". Neuroimaging had confirmed a lot of what was known from brain mapping studies, he says, but it was not doing anything new. "We would just tweak a psych test and see what happens," says Owen. As for real clinical applications: "I realized there weren't any. We all realized that."

Owen wanted to find one. He and his colleagues got their chance in 1997, with a 26-year-old patient named Kate Bainbridge. A viral infection had put her in a coma — a condition that generally persists for two to four weeks, after which patients die, recover fully or, in rare cases, slip into a vegetative or a minimally conscious state — a more recently defined category characterized by intermittent hints of conscious activity.

Months after her infection cleared, Bainbridge was diagnosed as being in a vegetative state. Owen had been using positron-emission tomography in healthy people to show that a part of the brain called the fusiform face area (FFA) is activated when people see a familiar face. When the team showed Bainbridge familiar faces and scanned her brain, "it lit up like a Christmas tree, especially the FFA", says Owen. "That was the beginning of everything." Bainbridge was found to have significant brain function and responded well to rehabilitation[3]. In 2010, still in a wheelchair but otherwise active, she wrote to thank Owen for the brain scan. "It scares me to think of what might have happened to me if I had not had mine," she wrote. "It was like magic, it found me."

Owen moved from visual to auditory tests — "up the cognition ladder, from basic sound perception, to speech perception and then to speech comprehension". For example, he presented people in a vegetative state with phrases containing words that sound the same but have two meanings, such as "The dates and pears are in the bowl". The ambiguity forces the brain to work harder and shows up in characteristic fMRI patterns in healthy people — if, that is, they are comprehending the words. One of Owen's patients, a 30-year-old man who had been incapacitated by a stroke, showed the same pattern[4]. But not everyone was convinced that these signs pointed to comprehension. "Every time I would go to a neurologist or anaesthesiologist and say, 'he's perceiving speech', they'd ask 'but is he conscious?'." Owen realized that he needed a different experiment to persuade the sceptics.

ANYONE FOR TENNIS?

It was June 2006. Wimbledon was on, and in a headline-stealing study, Owen took fMRI scans of a 23-year-old woman in a vegetative state while he asked her to imagine playing tennis and walking through the rooms of her house. When healthy, conscious adults imagine playing tennis, they consistently show activation in a region of the motor cortex called the supplementary motor area, and when they think about navigating through a house, they generate activity in the parahippocampal gyrus, right in the centre of the brain. The woman, who had been unresponsive for five months after a traffic accident, had strikingly similar brain activation patterns to healthy volunteers who were imagining these activities, proving, in Owen's mind, that she was conscious. The result, published in a one-page article in *Science*[5], evoked wonder and disbelief. "I got two types of e-mail. People either said 'this is great' or 'how could you possibly say this woman is conscious?'," Owen says.

Other researchers contended that the response was not a sign of consciousness, but something involuntary, like a knee-jerk reflex. Daniel Greenberg, a psychologist at the University of California, Los Angeles, suggested in a letter to *Science* that "the brain activity was unconsciously triggered by the last word of the instructions, which always referred to the item to be imagined"[6].

But Owen went on to bolster his case. Working with neurologist and neuroscientist Steven Laureys from the University of Liège, Owen showed that of 54 patients in a vegetative or minimally conscious state, five responded in the same way as the first woman[1]. Four of them were in a vegetative state. After refining their methods, the researchers asked

JOHN HRYNIUK

patient 23 to use that capability to answer yes-or-no questions: imagine playing tennis for yes, navigating the house for no. They then asked about things that the technicians scoring the brain scans couldn't possibly know.

Is your father's name Thomas? No. Is your father's name Alexander? Yes. Do you have any brothers? Yes. Do you have any sisters? No. The experiment is no easy feat for the patient. Owen's protocol demands patients maintain focus for 30 seconds then rest for 30 seconds, with lots of repetition.

In front of a computer screen showing the fMRI data, Owen traces a blue line indicating activity in the supplementary motor area — a 'yes' — as it rises during the 'answer' period. It dives during the rest periods. A red line — indicating activity in the parahippocampal gyrus — represents the 'no'. The lines are sharp and clear, and Owen, who has a taste for puns, calls the implication "a no-brainer". "You don't need to be a functional-imaging expert to appreciate what this person is telling you," he says. The patient answered five of six questions correctly[1]. There was no discernible signal for the sixth.

"IT WAS LIKE MAGIC. THE BRAIN SCAN FOUND ME."

Russell Poldrack, a neuroimaging expert at the University of Texas at Austin, calls Owen's methods ingenious. "When I want to give someone examples in which fMRI has told us something we really didn't know before, I use these," he says.

But Parashkev Nachev, a clinical neuroscientist at Imperial College London, criticizes the work for "assuming that consciousness is a binary phenomenon". Many patients, such as those having certain types of epileptic seizures, exhibit limited responsiveness without being conscious. Nachev says that more data are needed to indicate where in the continuum of cognitive abilities people in vegetative states fall.

Owen agrees that consciousness is not an "on-or-off thing". He sees it as an "emergent property" of many "modules" of the brain working together. Enough of these modules are at work in his exercise, he says, for responsive patients to qualify as being conscious. A person needs long-term memory to know what tennis is, short-term memory to remember the question or command and intention to give an answer. Ultimately, Owen is not concerned with pinpointing a threshold of consciousness or with providing a comprehensive definition for it. He takes a "know it if you see it" approach. Responding to commands and questions — communication — is an undeniably conscious activity, in his view. "In the end if they say they have no reason to believe the patient is conscious, I say 'fine, but I have no reason to believe you are either'," he says.

TO THE CLINIC

Currently, there are tens of thousands of people in a vegetative state in the United States alone. Owen reckons that up to 20% of them are capable of communicating; they just don't have a way to do so. "What we're seeing here is a population of totally locked-in patients," Owen says.

Owen now wants to put his technique into the hands of clinicians and family members. So far, the technology has done little. The first woman in the tennis study died last year, and patient 23, for logistic and financial reasons, was assessed only once. Even if a person in a vegetative state is 'found', there is no guarantee that he or she will later be able to return a normal life. Owen nevertheless insists that "clarifying" a patient's state of consciousness helps families to deal with the tragedy. "They want to know what the diagnosis really is so that they can move on and deal with that. Doubt and uncertainty are always bad things."

Two years ago, Owen was awarded a 7-year Can$10-million Canada Excellence Research Chair and another $10 million from the University of Western Ontario. He is pressing forward with the help of three new faculty members and a troop of postdocs and graduate students.

An early goal of the programme was to repeat the fMRI findings using an electroencephalogram (EEG)[7]. An EEG lacks fMRI's precision, and it cannot look as deeply into the brain, so the regions active in the tennis study were "off the menu", says Owen. But other tasks — imagining wiggling a finger or toe — produce signals that, through repetition, become clear. An EEG is also cheap, relatively portable and fast (with milliseconds of lag compared with 8 seconds for fMRI), meaning that the research team can ask up to 200 questions in 30 minutes. "From a single trial you're not going to say, 'that person is saying yes', but if they get 175 of 190 right when tested, it's pretty clear."

Now, using an EEG, Owen is planning to study 25 people in a vegetative state every year. He will have the help of a new 'EEGeep', a jeep equipped with experimental equipment that will allow the researchers to travel around to test patients who cannot be transported to Western Ontario.

One goal is to identify other brain systems, such as smell or taste, that might be intact and usable for communication. Imagining sucking a lemon, for example, can produce a pH-level change in the mouth and a recognizable brain signal[8]. Owen has shown that registering jokes provokes a characteristic response in healthy people[9] and plans to try it on patients in a vegetative state. He hopes that he can use these tests to find some level of responsiveness in patients who cannot produce the tennis and navigation patterns of activity because of their level of brain damage.

The studies will also explore whether these patients have the capacity for greater intellectual depth. Owen thinks that some people in a vegetative state will eventually be able to express hopes and desires, perhaps like French magazine editor Jean-Dominique Bauby, who dictated his memoirs by repeatedly winking one eye. "I don't see a reason why they could not have a similar richness of thought, although undoubtedly some will not," Owen says.

His techniques could also radically change treatment. Owen is already asking patients whether they feel pain. The answers will be useful in dosing pain killers, and similar tests could even be used in intensive-care units to guide rehabilitation resources, says Loretta Norton, a graduate student who is undertaking a study for this purpose. But she recognizes that this will be controversial.

DECISION TIME

Owen's methods raise more difficult dilemmas. One is whether they should influence a family's or clinician's decision to end a life. If a patient answers questions and demonstrates some form of consciousness, he or she moves from the 'possibly allowed to die' category to the 'not generally allowed to die' category, says Owens. Nachev says that claiming consciousness for these patients puts families in an awkward position. Some will be given hope and solace that their relative is still 'in there somewhere'. Others will be burdened by the prospect of keeping them alive on the basis of what might be ambiguous signs of communication.

Even more ethically fraught is whether the question should be put to the patients themselves. Fins and Schiff question whether patients would ever be able to show that they can understand the complexities of that question in the way that is normally demanded of, for example, patients giving informed consent.

Owen hopes one day to ask patients that most difficult of questions, but says that new ethical and legal frameworks will be needed. And it will be many years, he says, "before one could be sure that the patient retained the necessary cognitive and emotional capacity to make such a complex decision". So far, he has stayed away from the issue. "It might be a little reassuring if the answer was 'no' but you can't presuppose that." A 'yes' would be upsetting, confusing and controversial.

For now, Owen is hoping to use the technology to find other responders like Kate Bainbridge — who Owen now describes as a "motivational force". "Otherwise," he says, "what's the point?" ■

David Cyranoski *is* Nature*'s Asia-Pacific correspondent.*

1. Monti, M. M. *et al. N. Engl. J. Med.* **362,** 579–589 (2010).
2. Fins, J. J. & Schiff, N. D. *Hastings Center Report* **40,** 21–23 (2010).
3. Menon, D. K. *et al. Lancet* **352,** 200 (1998).
4. Owen, A. M. *et al. Neuropsychol. Rehabil.* **15,** 290–30 (2005).
5. Owen, A. M. *et al. Science* **313,** 1402 (2006).
6. Greenberg, D. L. *Science* **315,** 1221 (2007).
7. Cruse, D. *et al. Lancet* **378,** 2088–2094 (2011).
8. Wilhelm, B., Jordan, M. & Birbaumer, N. *Neurology* **67,** 534–535 (2006).
9. Bekinschtein, T. A., Davis, M. H., Rodd, J. M. & Owen, A. M. *J. Neurosci.* **31,** 9665–9671 (2011).

COMMENT

CHANDRA X-RAY OBSERVATORY CENTER

The Milky Way as a combined image of near-infrared (yellow), infrared (red) and X-ray (blue and violet) data collected by NASA's 'Great Observatories'.

A midlife crisis for X-ray astronomy

As the field celebrates its 50th birthday, **Martin Elvis** asks how to keep this unique window into the Universe open.

On 18 June 1962, an 8-metre rocket carried three small X-ray detectors to the edge of space. They spent just under 6 minutes above the altitude of 80 kilometres, high enough for kiloelectronvolt-energy X-rays from space to reach them through the thinned atmosphere. The result of this brief flight by physicist Riccardo Giacconi and his colleagues revolutionized astronomers' view of what the Universe contains.

For 50 years, X-ray astronomy has burgeoned. The field has grown a billion times more sensitive — a feat that took optical astronomy 400 years to achieve — and has opened up such exotica as black holes and dark matter to detailed investigation. It has provided a unique window into 'extreme' places of the Universe, where gas can be 1,000–10,000 times hotter than the surface of the Sun. The question now is whether that progress will continue.

In the 1960s, X-rays were known to be produced by the Sun — that had been discovered in 1948. But the Sun's weak X-ray output made it seem futile to try to detect any other star this way. Giacconi justified his flight by telling the US Air Force that he was looking for X-rays from the Moon, but he hoped to find more. Fortunately, the Universe delivered surprises: a strong source of X-rays in the constellation Scorpius and a bright background of X-rays from all over the sky. The researchers were lucky. Had they looked when Sco X-1, as this first source was dubbed, was below the horizon, they would have detected only the bright background noise. Interest in cosmic X-rays might have withered.

Many teams rushed to follow up in the subsequent years. X-ray astronomy papers often dominated the pages of *Astrophysical Journal Letters*. One important discovery was that the brightness of these cosmic X-ray sources often changed within seconds. Because nothing can change its ▶

output faster than light can travel across it, this meant that the objects must be tiny — just light seconds across. Creating so much energy from such a small volume required some not-yet recognized energy source. This turned out to be gravity: the accretion of gas falling down the deep gravitational well of a compact star converts potential energy into millions of degrees of heat.

The next big step was the development of orbiting X-ray instruments, which allowed for exposure times of days rather than minutes. NASA's 1970 Uhuru mission, led by Giacconi, was the first in a small fleet of these. Uhuru and its successors detected hundreds of X-ray sources 1,000–10,000 times fainter than Sco X-1.

But it was the cosmic X-ray background that set the real programme for X-ray astronomy for the next 40 years. Was it caused by hot gas pervading intergalactic space, or by millions of faint and distant sources blending together? Early detectors could not answer this question, because they had to collect X-rays from a large chunk of the sky to detect a signal; discrete sources that might make up the background produced no more than one X-ray per square centimetre of detector each day. To see with greater precision, researchers needed more sensitive telescopes, which they built from nested cylindrical mirrors.

The first such imaging X-ray telescope — NASA's 1978 Einstein Observatory, led by Giacconi — detected radiation 100-fold fainter than any of its predecessors. It found that 20% of the cosmic X-ray background came from active galactic nuclei, later understood to host supermassive black holes in their centres.

LEAPS AND BOUNDS

By 1987 — the twenty-fifth anniversary of the discovery of Sco X-1 — X-ray astronomy had become 10 million times more sensitive. The next generation of X-ray telescopes came in the early 1990s. Notable among them was Germany's ROSAT, which surveyed the whole sky, cataloguing 100,000 X-ray sources and, by going tenfold fainter still than Einstein, resolving some 60% of the cosmic X-ray background that had an energy lower than 2 kiloelectronvolts (keV).

The current era was ushered in by the launch of two large X-ray observatories in 1999: NASA's Chandra and the European Space Agency's XMM-Newton. XMM has a large collecting area, but Chandra — with an angular resolution ten times better than anything previous — was revolutionary. Giacconi received the Nobel prize in 2002, perhaps stimulated by the outpouring of results from Chandra, which he had been instrumental in getting started.

Chandra has resolved the cosmic X-ray background at all energies up to 7 keV. More than 90% of the background is due to the summed emissions of millions of active galaxies (most of the rest comes from fainter, star-forming galaxies). Chandra and XMM have also delimited the extent of dark matter's interaction with itself, determined properties of dark energy and showed that as much as one-fourth of the energy of a supernova goes into accelerating protons, solving the mystery of the origin of high-energy cosmic rays.

Fifty years on, the founding mysteries of X-ray astronomy have been solved. But our understanding of more complex questions is still primitive. In my own sub-field alone, we don't yet know why active galaxies emit X-rays, where the massive black holes at their centres come from or how a massive black hole accelerates matter in bulk to speeds more than 90% that of light in a tight beam that can extend over millions of light years.

> ***"Prospects for a set of 'Greater Observatories' that span the electromagnetic spectrum look bleak."***

At the same time, X-ray astronomy has the power to test two bases of twentieth-century physics — general relativity and quantum chromodynamics (which describes interactions of the 'strong force' between quarks and gluons) — in the extreme conditions around black holes and neutron stars.

But for astronomers to answer these questions, X-ray astronomy must become even more sensitive, by collecting more photons with larger mirrors while maintaining the fine resolution seen with Chandra. The US 2010 decadal survey for astronomy ranked a larger-area X-ray mission fourth among major space missions, but that is low enough, given the restricted US science budget, to put a new X-ray observatory on the back burner for at least a decade. Each space observatory, be it optical, infrared or X-ray, now costs more than US$2 billion. The worldwide space astronomy budget of about $5 billion a year is not enough to sustain a comprehensive programme.

NASA's current 'Great Observatories' — Hubble, Chandra and Spitzer — span the infrared to X-ray bands. They will probably have just one successor: the James Webb Space Telescope (JWST), which was designed to study cosmology and galaxy evolution in the distant Universe, and so works primarily in the infrared. Over the rest of the spectrum, astronomers will be blind. A US successor to Chandra is unlikely until 2030. By then, the JWST will probably be dead, preventing cross-fertilization of ideas among different wavelengths.

> **NATURE.COM**
> For more on the benefits of private space enterprises:
> go.nature.com/4288nl

Ingenious, but more specialized, missions in X-ray astronomy can have more modest price tags of around $200 million. In the next few years, several such missions will be launched — one in pursuit of higher-energy rays (NuSTAR, part of NASA's Explorer class), one as an updated ROSAT (Germany's eROSITA) and one with relatively high spectral resolution (Japan's ASTRO-H). But none of these even approaches Chandra's spatial resolution.

Since 2010, some ideas for achieving more powerful missions on a tight budget have emerged, including combining the fine imaging of Chandra with a light-gathering surface 30 times more powerful. This would require the development of mirrors that can be actively corrected to ensure a sharp image, which seems feasible. But such a project would cost at least $2 billion. Prospects for a set of 'Greater Observatories' that span the electromagnetic spectrum look bleak.

A PROFITABLE SOLUTION

The only viable solution is to lower the cost of getting equipment into orbit. Launch costs have held steady at some $10,000 per kilogram of payload for more than 50 years. Spacecraft must have ingenious designs to keep their masses low, which makes the cost of building and launching high: $100,000 or more per kilogram of craft. Overall mission costs could plummet if launch costs were to decline.

The best — perhaps only — way to lower launch costs is to allow private enterprises to profit from space ventures. Profit is the counterweight to caution, and competition will drive down costs. Several companies have already developed operational launchers for government payloads, including Orbital Sciences Corporation of Dulles, Virginia, whose Pegasus rocket was, as *Nature* went to press, scheduled to launch NuSTAR on 13 June, and SpaceX of Hawthorne, California, which last month launched the first private mission to resupply the International Space Station.

Ultimately, a healthy market that spurs lower costs will require customers other than the government, such as those seeking minable resources more plentiful in space than on Earth. One new company, Planetary Resources of Bellevue, Washington, announced in April its aim to mine asteroids.

Such profit-seeking could decrease launch costs by more than an order of magnitude, such that fuel costs dominate the overall price. Then, and only then, will new generations of greater observatories, in X-rays and across the electromagnetic spectrum, become affordable for scientists. ■

Martin Elvis *is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.*
e-mail: elvis@cfa.harvard.edu

D. PIGNATELLI/BLOOMBERG VIA GETTY

Water levels in the Mekong Basin could rise or fall with climate change — models cannot say which.

Climate models at their limit?

Estimates of climate-change impacts will get less, rather than more, certain. But this should not excuse inaction, say **Mark Maslin** and **Patrick Austin**.

For the fifth major assessment of climate science by the Intergovernmental Panel on Climate Change (IPCC), due to be released next year, climate scientists face a serious public-image problem. The climate models they are now working with, which make use of significant improvements in our understanding of complex climate processes, are likely to produce wider rather than smaller ranges of uncertainty in their predictions. To the public and to policymakers, this will look as though the scientific understanding of climate change is becoming less, rather than more, clear.

Scientists need to decide how to explain this effect. Above all, the public and policymakers need to be made to understand that climate models may have reached their limit. They must stop waiting for further certainty or persuasion, and simply act.

Why do models have a limited capability to predict the future? First of all, they are not reality. This is perhaps an obvious point, but it is regularly ignored. By their very nature, models cannot capture all the factors involved in a natural system, and those that they do capture are often incompletely understood. Science historian Naomi Oreskes of the University of California, San Diego, and her colleagues have argued convincingly that this makes climate models impossible to truly verify or validate[1].

The more-concrete, less-philosophical problems can be illustrated by following the path of cascading uncertainties that are building up in the models used today.

One of the first inputs into any climate model is the expected accumulation of greenhouse gases and aerosols in the atmosphere by the end of the century. These projections are based on economic models that predict global fossil-fuel use over 100 years given broad assumptions about how green the global economy will become. The economic collapse of 2008 showed dramatically, and to our cost, how difficult it is to predict changes in the economy. And economic unpredictability is just the beginning.

Another layer of uncertainty comes from how the global climate models are weighted. For example, in the most recent IPCC assessment, released in 2007, the economic scenarios were input into more than 20 general circulation models. Every model has its own design and parameterizations of key processes, such as how to include the effects of clouds; and every model and its output was assumed to be equally valid, even though some perform better than others in certain ways when tested against historic records. The differences between the models will be exacerbated in the 2013 IPCC assessment, because many, but not all, of the models have improved spatial resolution.

The outputs from the circulation models are then often used to drive detailed regional climate models to predict local environmental variations. Such regional models have huge uncertainties, thanks largely to the fact that precipitation is highly variable over small scales of time and space. This leads to a large range of potential futures, some of which contradict others. For example, detailed hydrological modelling of the Mekong River Basin using climate model input from the UK Met Office's HadCM3 model projects changes in annual river discharge that range from a decrease of 5.4% to an increase of 4.5% (ref. 2). Changes in predicted monthly discharge are even more dramatic, ranging from a fall of 16% to a rise of 55%. Advising policymakers becomes extremely difficult when models cannot predict even whether a river catchment system will have more or less water.

Projected regional changes are then used as a basis for 'impact models' that estimate the effect on the quality of human life. But these effects often depend more on the relative resilience of a given society than on the magnitude of environmental change. Even the most advanced socioeconomic models, which look at the monetary costs arising both in market and non-market sectors, often fail to account adequately for major aspects of human suffering that are hard to quantify[3].

STRUGGLE WITH EXTREMES

A key debate has emerged between scientists about how well models can predict extreme climates. On the optimistic side, Tim Lenton of the University of Exeter, UK, has argued that, with more research, models will help to provide an early warning system of climatic tipping points such as the melting of the Greenland ice sheet, the dieback of the Amazon rainforest and the shift of the West African monsoon[4]. By contrast, Paul Valdes of Bristol University, UK, argues that climate models are too stable, built to 'not fail' rather than to simulate abrupt climate change[5]. When the current IPCC models were tested against four major past climate changes, he notes, two were unable to even get the basic

COMMENT

climate before the shift correct and the other two had to be fed parameters up to ten times greater than would be realistic to produce the abrupt shift.

The climate models, or 'climate simulators' as some groups are now referring to them, being used in the IPCC's fifth assessment make fewer assumptions than those from the last assessment, and can quantify the uncertainty of the complex factors they include more accurately. Many of them contain interactive carbon cycles, better representations of aerosols and atmospheric chemistry and a small improvement in spatial resolution.

Yet embracing more-complex processes means adding in 'known unknowns', such as the rate at which ice falls through clouds, or the rate at which different types of land cover and the oceans absorb carbon dioxide. Preliminary analyses show that the new models produce a larger spread for the predicted average rise in global temperature. Additional uncertainty may come to light as these models continue to be put through their paces. Dan Rowlands of the University of Oxford, UK, and his colleagues have run one complex model through thousands of simulations, rather than the handful of runs that can usually be managed with available computing time. Although their average results matched well with IPCC projections, more extreme results, including warming of up to 4 °C by 2050, seemed just as likely[6]. As computing power becomes more accessible, that 'hidden' uncertainty will become even more obvious.

STABLE AND CONFIDENT

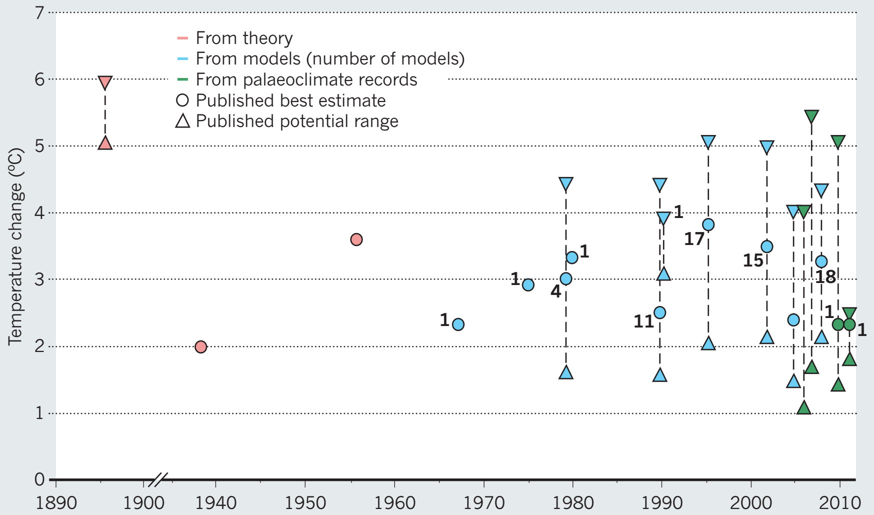
None of this means that climate models are useless. The present models are clearly able to reproduce natural climate variability over the past 150 years, and have provided an essential test of the theoretical link between CO_2 and global temperatures. Their vision of the future has in some ways been incredibly stable. For example, the predicted rise in global temperature for a doubling of CO_2 in the atmosphere hasn't changed much in more than 20 years (see 'Prediction stability').

This message of stability and confidence is often lost on the public: when Andreas Schmittner of Oregon State University in Corvallis published a value for the climate's sensitivity to a doubling of CO_2 that was on the low side of previous estimates[7], the media declared that "climate sensitivity was overestimated". This was despite some significant criticisms of the study and the fact that the findings still fell within the range of those in all the IPCC reports. This is a sad indicator of the headlines we might expect in the wake of the next IPCC report.

One approach to tackling the public-perception problem is to subtly rephrase the conclusions, placing the uncertainty on the

PREDICTION STABILITY

Estimates of climate sensitivity — the rise in global temperature caused by a doubling of atmospheric carbon dioxide levels — have remained fairly steady for decades.



References for data supplied in Supplementary information. Details can be found at go.nature.com/8dzwxv.

date by which things will happen, rather than onto whether they will happen at all. A recent study, for example, showed that the politically expedient 2 °C limit will be reached between 2040 and 2100, depending on our emission pathway and the model used[8]. This 'when' not 'if' approach is powerful.

In the face of scientific uncertainty, various philosophies for decision-making have arisen. But perhaps the best approach is to ensure that policies include 'win–win' strategies. Supporting a huge increase in renewable energy would reduce emissions and help to provide energy security by reducing reliance on imported oil, coal and gas. Reduced deforestation and reforestation should draw-down CO_2 from the atmosphere and help to retain biodiversity, stabilize soils and provide livelihoods for local people through carbon credits. Measures that lessen car use will increase walking and cycling, which in turn reduces obesity and heart attacks. No one can object to creating a better world, even if we turn out to be extremely lucky and the scale of climate change is at the low end of all projections.

"The biggest obstacle is the unwillingness of politicians to act in the long-term interests of society."

The biggest obstacle is the unwillingness of politicians to act in the long-term interests of society. Politicians use public opinion and scientific uncertainty as excuses for inaction. They used to say "we need to wait until scientists prove that mankind is causing climate change". That hurdle has, arguably, passed, so now they have moved on to "we need to wait until scientists can tell us exactly what will happen and what the costs are", or, "we need to wait for public opinion to be behind action". The former will never occur, because modelling can never provide that level of certainty. The latter is a sleight of hand. Politicians often take action without public support, from wars to bank bailouts, taxation to health-care reforms.

Greater knowledge and improved models will always be desirable, but they are not a panacea for political and public reticence to action on climate change. Despite the uncertainty, the weight of scientific evidence is enough to tell us what we need to know. We need governments to go ahead and act, as both the United Kingdom and Mexico have done in making national laws that contain carbon reduction targets of 80% and 50%, respectively, by 2050. We do not need to demand impossible levels of certainty from models to work towards a better, safer future. ■

Mark Maslin *and* **Patrick Austin** *are with the Environment Institute and Department of Geography, University College London, Pearson Building, Gower Street, London, WC1E 6BT.*
e-mail: m.maslin@ucl.ac.uk

1. Oreskes, N., Shrader-Frechette, K. & Belitz, K. *Science* **263,** 641–646 (1994).
2. Kingston, D. G., Thompson, J. R. & Kite, G. *Hydrol. Earth Syst. Sci.* **15,** 1459–1471 (2011).
3. Stern, N. *The Economics of Climate Change: The Stern Review* 692 (Cambridge Univ. Press, 2007).
4. Lenton, T. M. *Nature Clim. Change* **1,** 201–209 (2011).
5. Valdes, P. *Nature Geosci.* **4,** 414–416 (2011).
6. Rowlands, D. J. *et al. Nature Geosci.* **5,** 256–260 (2012).
7. Schmittner, A. *et al. Science* **334,** 1385–1388 (2011).
8. Joshi, M., Hawkins, E., Sutton, R., Lowe, J. & Frame, D. *Nature Clim. Change* **1,** 407–412 (2011).

Sanitation for all

Water pollution from sewage is causing great damage to India. The nation needs to complete its waste systems and reinvent toilet technologies, says **Sunita Narain**.

Sanitation is a much-sanitized word. It hides the horror of disease and the crippling indignity that people have to endure when they do not have access to a toilet. It also hides the technology divide for human excreta, which favours the rich in its collection, conveyance and disposal. The only solution is a complete sanitation system — toilets that are connected to a waste-removal and treatment system. But conventional technology does not work everywhere, or for all.

Rapidly-modernizing India is drowning in its own excreta. According to the World Health Organization, more than 600 million Indians practise open defecation — making up 60% of the 1.1 billion people who do so worldwide[1]. But even as toilets get built, the challenge of managing excreta grows.

More than 87% of people in India's cities (compared with 33% in rural areas) now have access to a toilet[2]. But leaking and incomplete sewage systems contaminate rivers and lakes, causing diseases such as cholera. Around 97 million Indians do not have access to clean drinking water, putting the nation second only to China[2]. Similar problems afflict other developing countries.

The challenge for India is to come up with ways of dealing with excreta that are affordable and sustainable. The first step is to match investment in waste-water systems with that for water supply. Innovative and affordable toilets must be designed and accessible to everyone. Only if all waste is treated can pollution be controlled.

THE SEWAGE SPIRAL

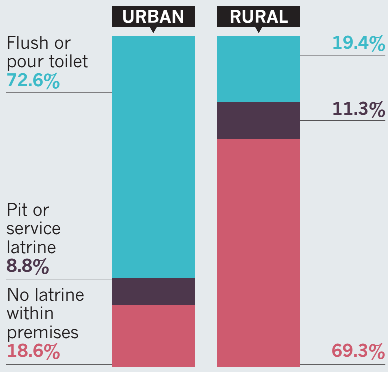
Emerging countries such as India are following a sanitation trajectory, gradually upgrading facilities from no toilets to sophisticated systems (see 'India's sanitation divide'). By 2011, just 20% of urban Indians had no toilet or used a public latrine[2]. But most of those with toilets had them connected to septic tanks or flowing into open drains. Only the top end — one-third of urban households and 2% of rural — had toilets that were connected to underground sewage networks, and not all of those reached treatment plants.

Progress along the trajectory is slow because the technology for collecting and disposing of excreta was invented in the water- and money-rich industrialized world. It is capital-intensive, favouring the rich over the poor. And it is resource intensive, using huge quantities of water.

City planners worry more about supplying water to their citizens than about the waste water generated. Yet the effluent inevitably goes into streams, lakes and rivers, or seeps into the ground to contaminate drinking water. Nitrate levels in groundwater across India exceed 45 milligrams per litre — a sure sign of sewage contamination[3].

INDIA'S SANITATION DIVIDE

City dwellers are more likely than people in rural areas to have flushing toilets, but only one-third of those toilets are connected to underground sewers; the rest go to septic tanks.



Leaked sewage leads to a deadly and costly spiral. As surface or groundwater gets contaminated, the city must source clean water from farther afield. The cost of pumping water rises — it now accounts for 30–50% of the price of supplying water. The cost of building and maintaining pipelines increases. And if the network is not maintained, water is lost — 30–50% of the water leaks. The nation therefore has less clean water to supply and needs to pay more to get it to the people. It cannot provide these services for everyone, and it chooses the rich. As the water system degrades, the rich move to bottled water. The poor get sick.

The over-burdened water utility then has no money to invest in sewage facilities. Most Indian cities do not have underground systems, and those that do, have old pipes that are in disrepair. There are few treatment plants. Officially, the country has the capacity to treat 30% of its waste water[4]. But in practice only 20% is processed: not all plants function and the pipes leak. The final blow comes when the treated waste water of the minority gets mixed with the untreated sewage of the majority.

Most believe that the sanitation divide will eventually disappear: one day, everywhere will have flush toilets, sewers and treatment plants. But in reality, Indian cities are way behind: they are growing so fast that their infrastructure cannot catch up. This cycle must be broken.

THINK OUTSIDE THE BOX

Governments must demand change in how water and waste are managed. They must cut the length of the pipeline, by investing in local water supplies such as lakes and ponds and using water more efficiently. They must invest more in sewage systems, even before they invest in water supply. Water accounts and tariffs must reflect the full cost of supply and of waste collection and treatment.

The challenge for science is to look beyond the modern engineering mindset. We know that current technologies, which use large amounts of clean water to transport small amounts of excreta through expensive pipes to costly treatment plants, are unworkable and unaffordable in much of the world. Yet toilets and sewage disposal are among the least researched technologies.

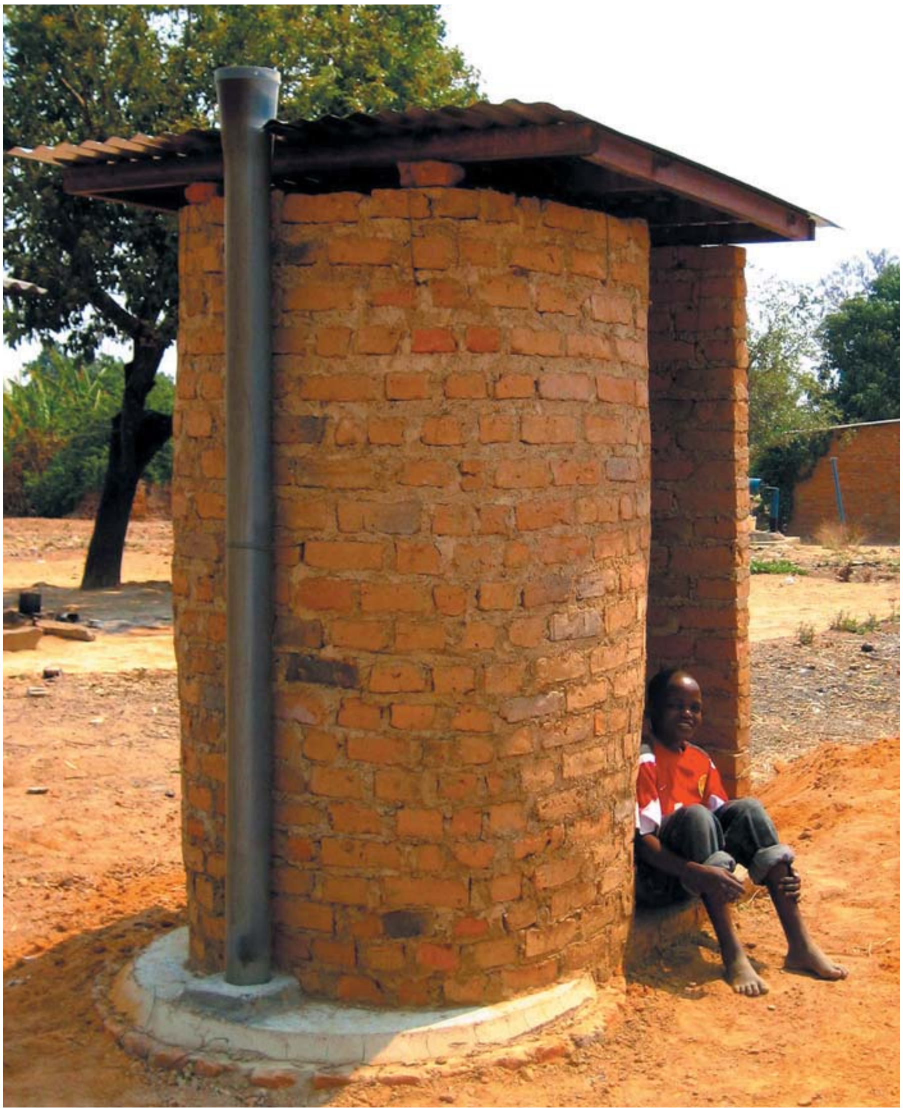
New technologies and new thinking are urgently needed for use across diverse ecosystems. Open drains might become planted waterways, with the vegetation cleansing the water. Or microbes might be used to decompose and de-pathogenize effluent. Sewage must be treated as a resource — turned into water for drinking, irrigation or industry.

Cities must leapfrog within the sanitation trajectory — go straight from no toilets to hygienic toilets for all that do not cost the Earth. Only then can the challenge be met. ■

Sunita Narain *is director general of the Centre for Science and Environment in New Delhi and author of* Excreta Matters: How Urban India is Soaking up Water, Polluting Rivers and Drowning in its Own Waste *(State of India's Environment, Centre for Science and Environment, 2012).*
e-mail: sunita@cseindia.org

1. WHO/UNICEF *Progress on Drinking Water and Sanitation* (WHO/UNICEF, 2012).
2. Office of the Registrar General and Census Commissioner *Census of India 2011* (Government of India, 2012).
3. Kamyotra, J. S. *Water quality and waste water management vision 2012–17* (Central Pollution Control Board, 2011).
4. CPCB *Status of Water Supply, Wastewater Generation and Treatment in Class-I cities and Class-II towns of India* (Central Pollution Control Board, 2009).

P. MORGAN



An 'eco' toilet such as Peter Morgan's Blair VIP, which traps disease-carrying flies, can transform lives.

ECODESIGN

The bottom line

If architecture is 'design for living', one of its greatest challenges is how to live with the masses of waste we excrete. Four pioneers in green sanitation design outline solutions to a dilemma too often shunted down the pan.

Every year, on average, each of us excretes 50 litres of faeces, rife with pathogens and heavy metals. Multiplied by Earth's population of 7 billion — and rising — that constitutes not so much an elephant in the room as a herd of mammoths. Sustainable solutions are urgently needed, particularly for the 2.6 billion people who lack adequate sanitation and the 1.1 billion practising open defecation. Rich countries, meanwhile, often have hidden sanitation issues of their own.

There is no single design solution to sanitation. But there are universal principles for systematically and safely detoxifying human excreta, without contaminating, wasting or even using water. Ecological sanitation design — which is focused on sustainability through reuse and recycling — offers workable solutions that are gaining footholds around the world, as *Nature* explores on the following pages through the work of Peter Morgan in Zimbabwe, Ralf Otterpohl and his team in Germany, Shunmuga Paramasivan in India, and Ed Harrington and his colleagues in California.

In compost-based ecosanitation, excreta is reframed as a resource: fertilizer. Much of the research on this has focused on upping nutrient levels and finding faster, more effective ways of removing heavy metals and pathogens such as viruses. Meanwhile, research-based, ecological processing of waste water is vastly improving water-based systems, bringing them closer to the ecosanitation ideal.

Environmental sustainability is only part of ecosanitation, however. Defecation is as culture-laden as other behaviours, so the designs must also be socially sustainable — tailored to local customs and strictures, whether in Malawi or Manhattan.

The developed world may think it has cracked the problem, but trouble is gurgling away underground. 'Flush and forget' sanitation systems constitute one of the more bizarre hangovers from the Victorian age. In older toilets, up to 25 litres of drinking water go down the pan per flush, although 'low-flow' toilet designs are coming into their own and, in 1995, the US federal government set a 7-litre-per-flush limit.

Aside from wasted water, the faeces-laden 'black water' from flush toilets is not always treated. Many older US and UK sewage systems, for instance, mix toilet waste water with storm water in so-called combined sewage outflows, which can overflow after heavy rain. The US Environmental Protection Agency estimated in a 2004 report to Congress that 850 billion gallons of untreated water were entering US waterways every year.

Sewage sludge — the semi-solid mush left after wastewater treatment in sewage works — can be as problematic. Although it can contain significant traces of pharmaceuticals and heavy metals even after treatment, it is widely used in the West as a soil conditioner and fertilizer on cropland, with uncertain effects on human health.

In the packed cities and scattered villages of the developing world, the challenge is even more daunting. Thousands of children die every day from a lack of basic sanitation or clean water. Open defecation contaminates soils with the eggs and larvae of soil-borne intestinal worms, or helminths, as well as other pathogens. More than one billion people are infected with these helminths, which cause, among other problems, weakness and malnutrition.

So a toilet can be transformative. A clean environment means better health — and that, in turn, is a springboard to development. As governments debate the finer points of global development challenges at Rio+20 next week, they might find it worth asking why sanitation falls to the bottom of most policy agendas. ■

PETER MORGAN

Inspired by ant turrets and the flight paths of flies

Environmental scientist and designer of the Blair VIP toilet, Harare

"The designer knows he has reached perfection, not when there is no longer anything to add, but when there is no longer anything to take away." I recalled this anonymous quote when, as a young biologist in Rhodesia (now Zimbabwe) in the 1970s, I was working for the health ministry's Blair Research Laboratory — named after former health secretary Dyson Blair. Blair had persuaded me to change fields, from schistosomiasis control to technological solutions for public health, and tasked me with designing new toilet systems for use in rural areas.

Open defecation was then common, and the existing pit toilets bred blowflies of the genus *Chrysomyia*, as well as other fly species that carry enteric disease. In a survey conducted between March 1974 and April 1975, we counted more than 20,000 flies emerging from a single pit toilet.

It had long seemed to me that simplicity could be related to elegance of design. My first innovation, called the Blair toilet after Dyson (and later dubbed the Blair ventilated improved pit, or BVIP toilet), is simple: a pit, lined with bricks for stability; a concrete sanitary slab with one hole for squatting, and another fitted with a vent pipe stretching from the slab to above roof level; and a spiral superstructure that obviates the need for a door but guarantees semi-darkness (see 'Air traffic control').

The design harnesses natural principles. The turrets of ant nests — the most elegant of nature's chimneys — inspired the vent. The natural behaviour of flies, which are attracted by odour and light, determined the other design features. When air passes over the top of the vent, suction draws more air through the squatting hole into the pit, then sends the odours up through the vent. Some flies are drawn to those odours; others, entering the pit through the squat hole, are drawn to the light from the bottom of the vent. Either way, the flies are trapped and die, because the vent is fitted with a non-corrodable screen, usually made of aluminium.

The concept is simple and it works. From October to December 1975, weekly counts of fly output were taken from two Blair toilets and two unventilated pit toilets: a total of 13,953 flies were trapped from the unventilated toilets and only 146 from the ventilated toilets (P. R. Morgan *Cent. Afr. J. Med.* **23,** 1–4; 1977). A family BVIP will last 10–15 years, and once it is abandoned, the superstructure materials can be recycled. The excreta gradually dries, and can be used as compost.

The BVIP is now the backbone of Zimbabwe's sanitation programme, with half a million family toilets built so far, and is widespread in other African countries. A multicompartment unit was designed for schools. More recently, I have drawn up a cheaper, upgradeable family unit, which can be built in stages — allowing them to ascend the 'sanitation ladder'.

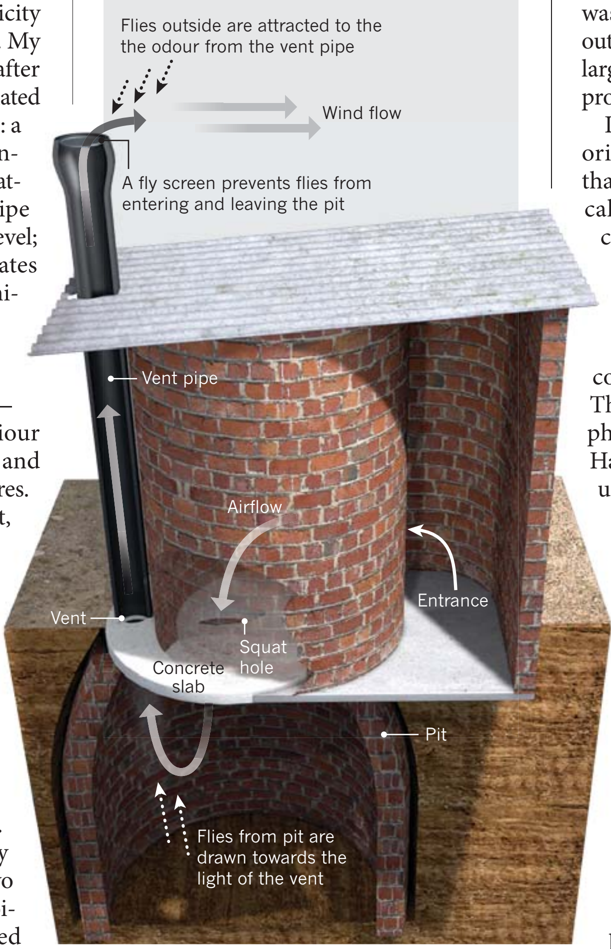
I devised other toilets to speed up the composting process. The Arborloo is an unlined pit 1 metre deep, which is fitted with a circular brick or concrete rim and a sanitary slab. As it fills, soil, ash and leaves are added to accelerate composting and control flies and odour. After a year, the toilet superstructure is moved to a new pit, and a tree is planted in a layer of soil on top of the old pit, to provide shade, fuel or fruit. Thousands of these have been built in Malawi and about 70,000 in Ethiopia.

The Fossa Alterna is another variation: two shallow pits are dug and used alternately, swapping at annual intervals. By the time one pit has filled, the compost in the other will be mature and ready for use.

Perhaps working in an area for which I have not received formal training has given me freedom of expression in observing, exploring and researching. I had to use instinct and plain logic. And it pays to adapt natural principles honed over millions of years.

AIR TRAFFIC CONTROL

The Blair toilet uses a vent to trap flies, preventing them from spreading disease. The vent also draws out odours from the pit, making the toilet interior almost odour free.



SOURCE: P. MORGAN

RALF OTTERPOHL

Boosting compost with biochar and bacteria

Environmental engineer and director, Institute of Wastewater Management and Water Protection, Hamburg University of Technology, Germany

With my team in Hamburg, I have studied resource-oriented sanitation — in which waste is seen as reusable — for 15 years. I was drawn towards this path having started out mathematically modelling mass flows at large-scale wastewater treatment plants — a process I found frustratingly inefficient.

In 2010, I began to focus on a practice originating in Brazilian Amazonia more than 1,000 years ago that could, paradoxically, kick-start a modern revolution in composting sanitation. The pre-Columbian Indians created 'black' soils known in Portuguese as *terra preta*. Found in patches throughout the Amazon, they are composed of charcoal (biochar), composted excreta and other bio-waste. They are absorptive and high in nitrogen, phosphorus, potassium and calcium. At Hamburg, we are adapting this mixture for use in ecosanitation systems.

Systems using terra preta technology can help to solve two problems that plague many developing countries: poor soils and a lack of sanitation. The technology offers the efficient creation of well-structured, humus-rich compost, which is important for food security, resistance to soil erosion, water retention in soil and the growth of local agricultural economies. And it is cost-effective. A basic terra preta sanitation toilet costs about US$50, inputs are cheap, and it is not hooked up to sewage systems.

Terra preta sanitation is a three-step process. First, lactic-acid bacteria are

There is a social side to sustainable sanitation: safe and attractive toilet blocks are being built for girls at schools in India, raising attendance levels.

WHEREVER THE NEED

added after each defecation: the anaerobic fermentation sanitizes and deodorizes. We have used cultures ranging from sauerkraut liquor to strains scientifically selected not to produce gas, such as *Lactobacillus plantarum*, *Lactobacillus casei* and *Pediococcus acidilactici*. At the same time, a waste-sugar source such as molasses or vegetable scraps is added as bacterial feed. This process continues while the excreta collect and for at least another week after collection ceases.

Second, about 50 grams of powdered charcoal — preferably 'clean', from wood-gas stoves — is added with each bacterial application to absorb odour and bind nutrients. This biochar also creates microporous space for the lactic-acid bacteria to inoculate the faeces and lower their pH value.

Finally, earthworms (vermiculture) compost the collected and treated material aerobically, further sanitizing the mixture over a period of three months to a year. The resulting compost is safe for use with non-food 'industrial' crops or forest trees. Two years of further processing creates compost that can be used on food crops.

Terra preta sanitation systems can be modified to suit rural or urban sites, but it must be possible to seal collection and transport receptacles for the fermentation to work. Rural versions can be simple bucket toilets fitted with a urine diversion, so that faeces drop into one bucket and urine is piped into another. Urine can be lacto-fermented separately with the bacterial mix for a month or more to avoid nitrogen losses and odours, then used as fertilizer, preferably on non-food crops. Alternatively, the urine can be used immediately after collection, mixed with at least five parts water.

For urban settings we are designing a new family toilet flushed with very small amounts of water, sprayed either manually with a bottle or through a nozzle attached to sink pipes. This toilet mixes faeces and urine. Because moisture impedes composting and the high nitrogen content must be compensated for with a carbon source, half a kilogram of woody waste must be added later at the composting site. This unit has a tank big enough for at least a week's worth of waste. The treated waste can then be collected and made into terra preta at a professional communal composting site.

Our research shows that these toilet systems and variations on the basic terra preta processes could work even in densely populated cities — a revolution indeed.

SHUNMUGA PARAMASIVAN

Creating social inclusivity for girls and women

Country director, Wherever the Need, Pondicherry, India

Since 2007, I've helped to design and install toilets across India that tackle specific environmental and social challenges — including those relating to women, schools and slums. All are environmentally sustainable urine-diversion dry (composting) toilets, which use little water and counter India's serious faecal pollution problem (see page 185).

The challenges we are up against are considerable. More than 626 million Indians practise open defecation daily, polluting groundwater and creating ideal breeding conditions for disease. Meanwhile, urban India's pace of growth is incredible — by 2030, estimates suggest that some 575 million people will be living in cities. One in four of the country's city dwellers (93 million) lives in slums.

But the social side of sustainability is as important as the environmental. Inadequate or no sanitation facilities in schools mean that once girls reach puberty they stay at home when they are menstruating, disrupting their studies. And women tend to excrete when it is dark because they are seeking privacy — but that makes them vulnerable to attack, or even rape. So our designs factor in safety and ease of use for girls and women.

The girls' school sanitary blocks we have designed in Cuddalore, Tamil Nadu, have an exclusive wash room with a washing facility, napkin vending machine and an incinerator. They are painted in brilliant colours, fully enclosed and imaginatively laid out so that they are enjoyable to use. We educate the girls with a software programme underlining the need for toilets, personal hygiene, and maintenance and management of the facility. Pupils play their part by helping to clean the toilets.

School sanitation is a large part of our focus, especially in rural villages. Although the older generation uses entrenched

methods of toileting, it is easier to convince children to use new ways. They then become agents of change for their families and communities.

In 2008, we installed 150 toilets in the slums of Cuddalore, which presented different issues. Family toilets with dedicated compartments to serve women and children are essential, but space was so limited we opted for mobile toilets. Two hundred families used the first unit; the second, redesigned to resemble a Portaloo, serves 280 families. We developed new single mobile ecosanitation units serving 15 to 20 families when the community asked for them. Mobile urinals were deployed in the most congested places, such as the hospital and harbour.

Both human compost and urine as a fertilizer are in high demand from farmers and Wherever the Need is now storing and treating both for eventual sale. The proceeds are paying for the logistics involved in this operation. Next we hope to replicate the operation in Chennai.

ED HARRINGTON
Building a coastal wetland in the heart of a city

General manager, San Francisco Public Utilities Commission, California, and former chair of the Water Utility Climate Alliance.

From the earliest planning stages, in 2009, for the new 13-storey headquarters of the San Francisco Public Utilities Commission, I knew the building would need to demonstrate the commission's ambitious sustainability goals — we needed to 'walk the talk'.

Water reuse was a central concern, but we quickly discovered that many decentralized water-reuse technologies were too energy-intensive for our building's energy budget. My team proposed that we pursue ecological sanitation methods inspired by wetlands, and challenge ourselves to defy conventional wisdom that the space such methods demand is too big for dense urban areas.

In our low-energy solution, wastewater treatment, which is usually buried in the basement, is visible — in the atrium and from the pavements where thousands walk every day. And we've done it in one of the highest-density neighbourhoods in one of the highest-density cities in the United States.

The building, which is expected to receive a leadership in energy and environmental design (LEED) platinum certification from the US Green Building Council in Washington DC, contains a sustainable, integrated water-management system that will serve more than 1,000 staff and visitors daily. The Living Machine system treats all waste water in the building for reuse in flushing toilets (see 'Water management wetlands style'). Rainwater is harvested from the roof and the child-care centre's play area for landscape irrigation. By treating and reusing all waste water on site in a self-contained system, we avoid discharging usable water.

Our system reduces total water use by about 65%, saving 3 million litres of drinking water per year. We anticipate eventually producing a surplus of 2.6 million litres of treated water for non-potable uses such as toilet flushing or park irrigation nearby.

The process begins and ends with low-flow toilets. After flushing, black water from the toilets is combined with grey water, from the sinks and showers, in plumbing pipes. The mixed waste water is then directed to tanks beneath the pavement for filtration; filtered solids such as faeces are pumped periodically to the municipal sewage system. From the tanks, the water is pumped into tidal-flow wetland 'cells', set next to the pavement and in the lobby, and disinfected using ultraviolet light and trace amounts of chlorine.

The cells are filled with gravel, bacteria and flowering grasses and plants — including Japanese sweet flag (*Acorus gramineus*) and western bleeding heart (*Dicentra formosa*). The plants were chosen for their root structure, which supports the establishment of treatment bacteria, and for resiliency in San Francisco's climate and urban environment.

The cells mimic the ecology and natural water-treatment processes of coastal wetlands, and are alternately flooded and drained to create multiple tidal cycles per day, speeding up the natural water-treatment processes by increasing the influx of oxygen. The system is responsible for recycling some 19,000 litres of water per weekday, but all people see is lush, vibrant plantings. It is a blend of function and aesthetics, as well as a demonstration of the ecological sanitation processes at work. ■

WATER MANAGEMENT WETLANDS STYLE

By carefully selecting plants that are suited to the local environment and can support particular bacterial communities, wetland cells provide an attractive way to treat and reuse all of a building's waste water on site.

INTERFACE MEDIA

OUP PAKISTAN

The underground sewage system that residents installed in Orangi has vastly improved public health.

Q&A Arif Hasan

Architect of change

Urban campaigner and architect Arif Hasan has been central to a sanitary revolution, transforming Orangi, Karachi, from informal settlement to thriving community. Using his technical know-how, residents built a sewage system, sparking vast social change. Now chair of Pakistan's urbanization task force, he discusses incorporating sustainable design into poor cities.

OUP PAKISTAN

What challenges did Orangi face when you arrived?

In 1980, the lanes of Orangi, a *katchi abadi* or informal settlement housing a million people, were running with waste water and sewage, and infant mortality was 128 in 1,000. The conditions stymied development: school attendance was down and trade difficult to establish. The psychological effects, too, were severe, sapping the will for change. The lanes couldn't be used as public space and quarrels over sanitation issues were frequent. The wastewater also damaged house foundations and triggered unhealthy rising damp.

How did you become involved?

That year, the social scientist Akhtar Hameed Khan set up the Orangi Pilot Project [OPP] to understand local problems and develop models to overcome the constraints governments face in upgrading informal settlements. Khan encouraged people to build and pay for their own underground sewage system, at a cost of around US$30 a household. A year later, he needed an architect. That was me.

How did you help?

I proposed cheap, simple, local solutions: concrete-curing methods, casting cheap steel-shuttering manhole covers *in situ*, site-appropriate tools such as manual compactors, and surveys and maps. I also developed rules of thumb for gradients, manhole and pipe sizes, survey procedures and inexpensive one-chamber septic tanks. These designs and methods, which challenged conventional engineering standards, have stood the test of time. They also reduced costs by more than 40%.

How did sanitation transform public health?

By 2000, some 85% of Orangi had self-laid, self-financed sewer lines. The lanes are clean: children play in them, women sit and talk. Health indicators have improved, and by 1993 infant mortality had fallen to 37 in 1,000.

What are the social and economic changes?

Literacy rates there are now among the highest in Pakistan. Socially and economically, Orangi is much more connected to the rest of Karachi in diverse ways. Many people are white-collar workers; a substantial minority are professionals; women and entrepreneurs work in and service formal-sector industry.

Is Orangi a model for other cities?

As an outgrowth of the process of building sanitation communally, residents became community activists and got involved in OPP-supported health, education, housing and micro-credit programmes. The architects and technicians who helped to initiate the project now run it, and have expanded it to cities and settlements all over Pakistan, often in collaboration with local governments and planning agencies. The OPP's philosophy has had a major impact on the attitudes of professionals, academic institutions, government officials and international non-governmental organizations and agencies working in Pakistan.

What other urban problems are you tackling?

In 1992, a square metre of land on the periphery of Karachi cost 1.7 times the daily wage. Today, it is 40 times. Many cannot afford that, so housing is getting denser. In the inner city, the density is more than 4,000 people per hectare [compared with less than 150 in London]. Although water supply and electricity provision have increased hugely in Karachi over the past 20 years, the sharing of toilets and kitchens has risen, along with the number of people per room. You can't live a comfortable life at such densities unless they are planned for.

What are your solutions?

With colleagues, I have developed design guidelines and incremental development processes to accommodate high densities in an environmentally friendly, affordable way. We hypothetically redesigned four different Karachi communities with computer modelling. Three redesigns showed that densities even higher than those prescribed by the Karachi Building Control Authority were workable.

What other challenges does Karachi face?

Many natural drainage channels and water bodies have been filled in to make way for elite coastal housing and new informal settlements upstream. So all of Karachi now floods. But other environmental issues are arising because of the new paradigm of Karachi as a 'world-class city', fuelled by direct foreign investment. New infrastructure caters to the corporate sector, and its architectural style imitates the West. There is a new dependence on air conditioning despite local expertise in creating buildings comfortable without it.

How has architecture in Karachi changed?

When I began my practice in 1968, an architect designed for the state, or for the rich. Today, clients vary hugely. This new architectural world belongs to women, who make up 92% of the architecture and planning students at the University of Karachi. In many ways, the role of the architect is immense compared to before. ■

INTERVIEW BY ANNA PETHERICK

Correspondence

Don't forget health in sustainability talks

Many wealthy nations are confronting the health and welfare implications of high unemployment, political instability and eroding social security. Such crises are part of our failure to move sustainability beyond rhetoric — a fact not adequately recognized by the Group of Eight (G8) community. The Rio+20 meeting in Brazil next week must accept that sustainable development and the health of populations depend on each other.

Peak global health may already be here. Low- and middle-income populations endure a double burden from communicable and non-communicable diseases, while health inequity and undernutrition persist. Social and ecological factors such as climate change, energy and food insecurity, counterfeit drugs, antimicrobial and insecticide resistance and poverty all undermine health in our interconnected world.

Many development specialists suggest that health has had its opportunity through the United Nations' Millennium Development Goals, and that the strategic imperative must shift to agriculture or energy. But these are only two legs of the stool: health remains the third.

When sustainability becomes a reality, it will reduce many risks to health. For example, clean energy will eliminate the need to clear forests and use crops for biofuel, so people will be better fed. Improved agricultural practices will reduce exposure to the animal-to-human infections that are associated with land clearing. A green revolution will protect biodiversity by producing more food on land that is already cleared. Better economic, educational and social opportunities in rural areas will slow migration and alleviate the pressure on overloaded urban health resources.

Bianca Brijnath *Monash University, Australia. bianca.brijnath@monash.edu*
Anthony J. McMichael, Colin D. Butler *National Centre for Epidemiology and Population Health, Australian National University, Australia.*

US policy should not shape collaborations

The aims of the Global Research Council (GRC) to improve international scientific collaboration (*Nature* **485,** 427; 2012) have been enthusiastically endorsed by John Holdren, director of the White House Office of Science and Technology Policy, and Subra Suresh, head of the US National Science Foundation (see go.nature.com/rbiykb). We disagree with their statement that "good science anywhere in the world is good for science — and good for people — everywhere in the world".

The agreed GRC standards for merit-based peer review will improve research quality through collaborations between high- and middle-income countries, which have strong research systems. But the guidelines are unlikely to boost science capacity in low-income countries.

Holdren and Suresh point out that US researchers could lose global funding if other governments do not review US proposals on their merit, and that the country's economic interests could be harmed if colleagues do not respect confidentiality and intellectual property. Subsequently, the heads of publicly funded science agencies from 47 countries agreed on the GRC's Statement of Principles for Scientific Merit Review.

It therefore seems that the GRC standards are transposing concern about intellectual property to the arena of peer review. This is reminiscent of the linking of intellectual-property rights with world trade law, which has created barriers to drug access in low-income countries.

Instead of implementing an aspect of US foreign policy that is intended to protect US economic interests, the GRC should promote scientific research in low-income countries. Apart from implying that they will discuss the expansion of open-access publishing in these countries, GRC members seem to have paid little attention to this issue.

Bridget Pratt, Bebe Loff *Monash University, Australia. bridget.pratt@monash.edu*

Planetary boundaries concept is valuable

As researchers studying the concept of setting environmental boundaries for planetary well-being (J. Rockström *et al. Ecol. Soc.* **14,** 32; 2009), we disagree with Simon Lewis's suggestion that there are flaws in these proposals (*Nature* **485,** 417; 2012).

Lewis contends that the concept's focus is too narrow, that it doesn't distinguish between 'boundaries' and 'thresholds' and that it should clarify the influence of scale on different problems. These are misunderstandings.

Planetary boundaries are not fixed 'supply limits', but are set within a safety margin around complex thresholds that are intertwined at regional and global scales. Ecosystem changes caused by nitrogen pollution, for example, are driven by global trade and cannot be uncoupled from climate change and alterations in land use. Also, investment in new phosphorus technologies can address the problems of both pollution and stock control.

It is a mistake to see the setting of environmental boundaries as a call for multilateral negotiations around static limits. They are instead a bid to reform environmental governance at multiple scales. Scientific assessment, institutional interaction and social–ecological innovation will enable us to work more effectively towards planetary stewardship.

Victor Galaz* *Stockholm Resilience Centre, Stockholm University, Sweden. victor.galaz@stockholmresilience.su.se*
**On behalf of 4 co-authors (for a full list, see go.nature.com/uct6dd).*

Combat the effects of Forest Code changes

On 25 May, Brazil's President Dilma Rousseff endorsed revisions to the country's Forest Code. Thanks partly to an appeal from the international scientific community and non-governmental organizations, together with a petition of more than 2 million signatures, Rousseff also vetoed some of the code's provisions. But more is needed.

The revised code will continue to protect forests in crucial locations, such as along rivers, on hilltops and in coastal wetlands, as well as a specified percentage of those on private property (see go.nature.com/fzxmj5). There will be no amnesty for offenders who illegally logged forests in the past.

I believe, however, that policies to reduce deforestation and foster sustainable development need to be strengthened to counter any potentially adverse effects of the new Forest Code. Payments for ecosystem services must be increased — for example, to expand the important work of the Forest Conservation Allowance Programme (Bolsa Floresta) in the Amazon. Ranching, agriculture and other economic activities should be confined to existing deforested areas. Environmental laws need stricter enforcement, and regular monitoring of forest areas is a priority.

Alison G. Nazareno *Federal University of Santa Catarina, Florianópolis, Brazil. alison_nazareno@yahoo.com.br*

NEWS & VIEWS

MICROBIOLOGY

Learning about who we are

Microbial inhabitants outnumber our body's own cells by about ten to one. These residents have become the subject of intensive research, which is beginning to elucidate their roles in health and disease. SEE ARTICLES P.207 & P.215

DAVID A. RELMAN

The dawn of the twenty-first century has seen the emergence of a major theme in biomedical research: the molecular and genetic basis of what it is to be human. Surprisingly, it turns out that we owe much of our biology and our individuality to the microbes that live on and in our bodies — a realization that promises to radically alter the principles and practice of medicine, public health and basic science. It is therefore appropriate that ever more research is focused on these microbes and their genes, which together are known as the human microbiome[1]. In this issue, the Human Microbiome Project Consortium[2,3] publishes the most extensive catalogue yet of organisms and genes pertaining to our microbiomes.

The first observations of indigenous human microbiota were published more than 300 years ago, soon after the invention of the microscope. Today's view of the microbial world has been radically improved by DNA-sequencing technology. In the wake of the Human Genome Project, calls were issued[1,4] for enhanced efforts to be made to characterize the 'second human genome' — the human microbiome. At the end of 2007, the US National Institutes of Health (NIH) launched the Human Microbiome Project (HMP) and, in early 2008, the European Commission and China initiated the Metagenomics of the Human Intestinal Tract (MetaHIT) project. Other countries have begun similar ventures, motivated in part by an interest in better defining their biological heritage.

Two studies, by Huttenhower *et al.*[2] (page 207) and Methé *et al.*[3] (page 215), together with 15 other papers[5,6] that are being published simultaneously elsewhere, comprise the first reports of the HMP Consortium research groups. The primary data, as described by Methé and colleagues[3], were derived from samples collected from 242 healthy adults in the United States, at 15 (for males) or 18 (for females) body sites — from the skin, nose, mouth, throat, vagina and faeces (to represent the distal gastrointestinal tract). Each person was sampled up to three times over 22 months, generating a total of 11,174 samples.

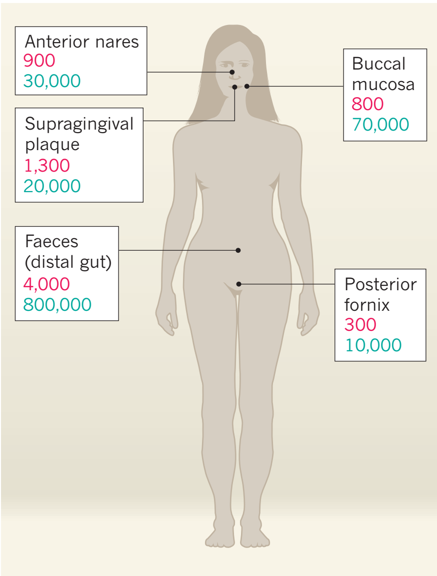


Figure 1 | Variation in diversity. Researchers of the Human Microbiome Project are studying the microbial inhabitants of the human body, using samples taken from 242 healthy adults at 15 (for males) or 18 (for females) body sites — from the skin (four sites), mouth and throat (nine sites), vagina (three sites), nostrils and faeces (to represent the distal gastrointestinal tract). Huttenhower *et al.*[2] and Methé *et al.*[3] have estimated the number of microbial species and their genes in these samples, and found substantial variation in microbial community composition at different body habitats. The two groups used different counting methodologies, and their numbers vary accordingly, such that exact figures are not available. However, crude estimations[3] of number of microbial species (red) and number of microbial genes (blue) are shown for examples of: sites containing high species diversity, such as the gastrointestinal tract and teeth (supragingival plaque); sites with intermediate diversity, such as the inside of the cheek (buccal mucosa) and nostrils (anterior nares); and sites with lower diversity, such as the vaginal posterior fornix. The authors also found substantial variation in both the diversity and the composition of the microbial communities at different sites within the same general body region.

The consortium researchers obtained the nucleotide sequence of the small-subunit ribosomal RNA — a molecule found in all cellular life — from microorganisms in 5,177 of these samples[3]. These sequences are commonly used to infer the genetic relationships between organisms. The researchers also surveyed the genomes of the microbes in 681 of the samples[3] using a shotgun sequencing approach, which generates random sequences (reads) from a complex pool of DNA molecules. The reads are then assembled on the basis of overlapping sequence similarity, allowing researchers to identify genes and to predict the functions of the proteins that they encode.

The investigators mapped their reads to all available microbial and viral genome sequences to assess community composition — the different types of microbes and their relative abundance — at the various body sites. The researchers also determined the whole-genome sequences of about 800 bacterial strains isolated from humans (from a planned total of 3,000); these sequences have been placed in public databases and can be used as reference genomes for comparative purposes. The consortium authors conclude[2,3] that they have identified the majority of the common microbial taxa and their genes present in these 242 healthy humans.

One of the great strengths of the HMP is that samples were collected simultaneously from multiple body habitats of the same individuals. This allowed Huttenhower *et al.*[2] to discover that taxonomic and genetic diversity were greatest in tooth and stool samples, intermediate in skin samples and on the inside surface of the cheek, and lowest in vaginal samples[2,3,7] (Fig. 1). The researchers report that each habitat is characterized by a small number of highly abundant 'signature' taxa, but that the relative representation of taxa and genes in each habitat varies considerably between individuals. In most samples, high-abundance taxa are accompanied by low-abundance taxa from the same genus, suggesting that within-community niche specialization occurs. These findings confirm those of an earlier study[8], which demonstrated that body habitat accounts for much of the variation in bacterial community composition. Although there is clear evidence for individuality in people's microbiome compositions, the limited temporal scope of the HMP data set prevents a robust analysis of how these communities change over time.

As shown previously for faecal samples[9],

the relative abundance of microbial genes associated with certain physiological pathways varied less between samples from the same habitat than did the relative abundance of taxa. This suggests that there is functional redundancy between microbial community members. The prevalence of low-abundance genes varied the most between habitats, and Huttenhower and colleagues speculate that the functions of these genes correspond to body-niche-specific activities. Interestingly, the researchers also found evidence of taxonomic co-variation across sites in an individual, such as between the communities of the skin and the saliva[2]. One possible explanation for this is that interactions between community members are subject to selection pressure by host-specific, and host-wide, environmental factors.

Studying the human microbiome has so far been a lesson in humility. Although the HMP and the MetaHIT project[10] are revealing vast amounts of previously uncharacterized microbial diversity within our 'home turf', the functions of these communities remain largely unknown. Moreover, Methé *et al.*[3] report that only 57% of the non-redundant gene families identified by the HMP and MetaHIT researchers were detected by both groups, with 34% of the gene families being detected only in the HMP data and 10% only by the MetaHIT project. This raises questions about the representativeness of the people in the two projects. In both cases, samples were taken from adults in developed nations who have relatively similar lifestyles and, in the case of the HMP, without inflammatory disease. Another recent report reveals[11] that populations living in less-developed regions of the world have markedly different microbiomes from those living in the United States.

It is also important to consider the definition of health. The most common cause for exclusion of people from the HMP was chronic gum disease, a condition that is increasingly regarded as 'normal' in developed countries. Furthermore, the prevalence of overweight and obese individuals continues to rise in many populations around the world, the chronic use of prescription drugs is becoming more common, urbanization is increasing, and our natural environment is changing in unexpected ways. Future studies of the human microbiome should accommodate such factors, which are likely to influence our microbial inhabitants.

Many areas of human-microbiome research warrant further investigation, but viruses and small non-bacterial organisms such as fungi deserve special attention, as do questions regarding the functions of the microbiome. We are essentially blind to many of the services that our microbial ecosystems provide — and on which our health depends[12] — and investigators desperately need new approaches for studying interactions between members of the microbial community and their human hosts. As Huttenhower and colleagues suggest[2], the fact that some of the microbiome's functions are likely to be performed by rare community members or to involve genes that are expressed at low levels will further complicate attempts to decipher their influence.

Despite the valuable initial findings from the HMP and other projects, multiple lines of enquiry remain. Which factors are responsible for the day-to-day or longer-term variation in the composition and functions of a person's microbiome? To what degree are such factors intrinsic to the microorganisms, related to the host, or, indeed, stochastic? Which mechanisms regulate bacterial colonization or invasion of the human microbiome, how does the microbiome respond to disturbance, and to what degree does this response involve the propagation of surviving organisms versus new colonization from outside? What is the basis of resilience in the human microbiome, and can it be predicted and restored? Such questions suggest that the work of Methé *et al.*, Huttenhower *et al.* and the numerous others studying the human microbiome is only just beginning. ■

David A. Relman *is in the Departments of Medicine and of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA, and at the Veterans Affairs Palo Alto Health Care System, Palo Alto, California.*
e-mail: relman@stanford.edu

1. Lederberg, J. *Science* **288,** 287–293 (2000).
2. The Human Microbiome Project Consortium *Nature* **486,** 207–214 (2012).
3. The Human Microbiome Project Consortium *Nature* **486,** 215–221 (2012).
4. Relman, D. A. & Falkow, S. *Trends Microbiol.* **9,** 206–208 (2001).
5. www.ploscollections.org/hmp
6. Segata, N. *et al. Nature Methods* http://dx.doi.org/10.1038/nmeth.2066 (2012).
7. Li, K., Bihan, M., Yooseph, S. & Methé, B. A. *PLoS ONE* **7,** e32118 (2012).
8. Costello, E. K. *et al. Science* **326,** 1694–1697 (2009).
9. Turnbaugh, P. J. *et al. Nature* **457,** 480–484 (2009).
10. Qin, J. *et al. Nature* **464,** 59–65 (2010).
11. Yatsunenko, T. *et al. Nature* **486,** 222–227 (2012).
12. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. M. & Relman D. A. *Science* **336,** 1255–1262 (2012).

QUANTUM PHYSICS

Majorana modes materialize

Elusive theoretical fantasies known as Majorana modes have been observed in a hybrid semiconductor–superconductor system. These emergent exotica open up promising prospects for quantum computation.

FRANK WILCZEK

The condensed-matter physics community has been galvanized by the apparent experimental discovery of Majorana modes, reported by Mourik *et al.*[1] in a paper published in *Science*. These entities, whose existence had been predicted[2] theoretically, could become major components in quantum engineering. Specifically, they might provide the basic elements — the qubits — for a quantum computer[3,4].

Over the past few decades, physicists have discovered that particle-like excitations called quasiparticles, found in condensed-matter systems, can have strange, fascinating and possibly useful properties, including fractional electric charge and unconventional quantum statistics. Majorana modes are a major addition to this universe of exotica.

To set the context and avoid misunderstanding, a brief reflection on the fundamental nature of quasiparticles is in order. The elementary building blocks of material systems are not negotiable; electrons, photons and atomic nuclei are what we get to work with. Furthermore, the basic interactions among those ingredients are known to high accuracy: Maxwell's electrodynamic and Schrödinger's quantum equations rule. So how, within that familiar and reliable framework, can new and exotic 'particles' arise?

Phonons are the original examples of quasiparticles. They were introduced[5] conceptually by Albert Einstein in 1907. Two years previously, he had proposed the idea that light has particle-like properties, being created and transmitted in discrete units, namely photons. Generalizing that intuition, he suggested that the vibrations of solids come in discrete packets, which we now call phonons. Einstein used this notion to explain an otherwise mysterious deficit of vibrational motion in diamond at low temperatures. Einstein's work pre-dated the modern understanding of solids (for example, the atomic nucleus was discovered only in 1911), but its central concepts endure. 'Holes' are another crucial kind of quasiparticle[6]. As their name suggests, holes represent the absence of an electron where

50 Years Ago

With the increasing use of irrigation in horticulture and agriculture and the developing interest in the effects of evapotranspiration, it is frequently necessary to locate the level of a shallow water-table. This can be done inexpensively and effectively, using a small-diameter observation well and an electric probe … Twelve observation wells and a probe … have been satisfactorily used by me, in soils ranging from fine clay to coarse gravel, for a period of eighteen months, in connexion with investigations of the water-table in the Thames flood-plain. It is almost certain, however, that, if necessary, these instruments would give trouble-free service for a much longer period.

From *Nature* 16 June 1962

100 Years Ago

A Reuter message from New York reports that a steamer arrived at Seward (Alaska) on Sunday covered with volcanic dust from an eruption at Katmai, in the Aleutian Islands. It is stated by those on board that a steady stream of volcanic fragments and ash followed a terrific explosion, spreading over the countryside. The sun was obscured. Although the vessel was seventy miles distant, at four o'clock on Thursday afternoon complete darkness set in and ash fell in a thick layer on the decks. It is estimated that volcanic ash covers three hundred square miles of fertile country. According to a telegram from Seattle (Washington State) the volcanic disturbance is rendering wireless telegraphic communication with Kadiak, Rospberry, and Afognac, three of the most important islands of south-western Alaska, impossible.

From *Nature* 13 June 1912

one would normally be expected. They come in discrete, localized units and move as if they are particles.

Those two classic examples are good illustrations of noteworthy peculiarities of quasiparticles. Phonons, although they arise from the correlated motion of atomic nuclei, are particle-like objects that in no way resemble atomic nuclei. Holes, despite their origin in the correlated motion of electrons, have very different properties from electrons; indeed, they have the opposite electric charge. Quasiparticles transcend the elementary units from which they are built. They are emergent objects, embodying recognizable — that is, reproducible and long-lived — organizations of energy within structured materials.

Different materials can support different kinds of quasiparticles. Of course, each material's quasiparticles depend on the structure of the material and exist only within that material. Modern physicists can exploit their broad understanding of matter to design materials whose quasiparticles will have interesting or useful properties. Transistors, the building blocks of modern electronics, are an outstanding, but by no means singular, success story. They orchestrate the motion of electrons and holes to amplify and switch electronic signals.

Majorana modes add a striking new variation to the quasiparticle theme[7]. Appearing as solutions to equations of a type invented[8] by Ettore Majorana in 1937, Majorana modes represent forms of excitation predicted — and now observed — to be available to a few very special and specific kinds of quasiparticles. They were discovered as mathematical possibilities within quantum field theory[9], and were also found in theoretical models of possible exotic superconductors and of a particular form of the quantum Hall effect[10], which occurs in two-dimensional electron systems held at low temperatures and subjected to strong magnetic fields.

In a brilliant paper, Alexei Kitaev then showed[2] how Majorana modes could arise at the ends of superconducting wires. His reasoning was (relatively) simple and transparent, and is at the heart of most subsequent proposals for the realization of Majorana modes — including those that underlie the experiments of Mourik *et al.* and of other researchers[11,12] who have also reported evidence of the phenomenon in different but related systems.

To become acquainted with Majorana modes, it may be helpful to draw on their analogy to a more familiar property possessed by many particles and quasiparticles — their spin. For example, an electron at a given position can be in either of two states, with spin 'up' or 'down'. Spin provides new dynamic options for particles that have it; that is, the possibility of spin-dependent interactions, including interactions that change the spin's direction. Majorana modes generate, for quasiparticles that support them, a kind of emergent spin. It is convenient to have a word for the concept 'quasiparticle that supports Majorana modes', and I have suggested[13] 'mode-icule'. To describe the quantum state of a mode-icule, we must specify a two-component wavefunction, just as we do for a (spin-ful) electron.

Building on previous theoretical designs[14,15] for detecting Majorana modes, Mourik *et al.*[1] studied the electrical properties of indium antimonide nanowires lying atop a substrate made of a conventional superconducting material. Indium antimonide is a semiconductor that displays a strong quantum interaction called spin–orbit coupling: the motion of the electrons in the material is strongly coupled to their spin. Interaction with the substrate's electrons induces superconductivity in the wire's electrons as well, through a phenomenon known as the proximity effect. Mourik *et al.* observed that, in such a hybrid semiconductor–superconductor device, electrons can tunnel into and out of the wire with no change in energy, at an energy independent of applied voltage. This shows, quite directly, that there are unusual quantum states, associated with the ends of the wire, which have the properties predicted for Majorana 'spins'. The sensitivity of the strength of this tunnelling to the magnitude and direction of applied magnetic fields also matches theoretical expectations for the states associated with Majorana modes.

The analogy of Majorana modes to spin proves inadequate when we come to describe systems of several identical mode-icules. Crucially, the emergent 'Majorana' states attached to separate mode-icules, unlike ordinary spins, are not entirely independent. On the contrary, those emergent states are, unavoidably, highly entangled. As a result, the quantum state of $2n$ mode-icules is described by a 2^n-dimensional wavefunction, whereas $2n$ independent spins would require 2^{2n} dimensions[16]. This constraint arises because mode-icules are entities called non-Abelian anyons[17], and so have unconventional quantum statistics — as opposed to fermions (such as electrons) or bosons (such as photons).

The wavefunction's dimensionality, although drastically reduced by entanglement, remains exponentially large in n. Interchange of mode-icules is accompanied by the complicated, but perfectly predictable, evolution of their wavefunctions in a complex mathematical region known as Hilbert space. This evolution of wavefunctions in Hilbert space embodies the elegant principles of Clifford algebra[7,18].

Theorists have developed ingenious and ambitious proposals for exploiting mode-icules, with their controllable entanglement, to enable quantum computation[3,4]. Until recently, their visions have been, at best, loosely tethered to laboratory reality. To do useful computations, we will need to create many mode-icules, and to develop the ability to move their 'spins' around one another.

Arrays of superconducting wires of the type developed by Mourik and colleagues[1] could each support mode-icules at their ends. By linking them together, and applying suitable electromagnetic fields, it ought to be possible to substantiate the theorists' dreams. Much difficult terrain separates proof of mode-icules' existence from the promised land of complex, scalable mode-icule circuits, but a beachhead has been established. ■

Frank Wilczek *is at the Center for Theoretical Physics, Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.*
e-mail: wilczek@mit.edu

1. Mourik, V. *et al. Science* **336,** 1003–1007 (2012).
2. Kitaev, A. Preprint at http://arxiv.org/abs/cond-mat/0010440 (2000).
3. Kitaev, A. *Ann. Phys.* **321,** 2–111 (2006).
4. Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Das Sarma, S. *Rev. Mod. Phys.* **80,** 1083–1159 (2008).
5. Einstein, A. *Ann. Phys.* **22,** 180–190 (1907).
6. Shockley, W. *Electrons and Holes in Semiconductors With Application to Transistor Electronics* (Van Nostrand Reinhold, 1950).
7. Wilczek, F. *Nature Phys.* **5,** 614–618 (2009).
8. Majorana, E. *Nuovo Cimento* **5,** 171–184 (1937).
9. Jackiw, R. & Rossi, P. *Nucl. Phys. B* **190,** 681–691 (1981).
10. Read, N. & Green, D. *Phys. Rev. B* **61,** 10267–10297 (2000).
11. Williams, J. R. *et al.* Preprint at http://arxiv.org/abs/1202.2323 (2012).
12. Deng, M. T. *et al.* Preprint at http://arxiv.org/abs/1204.4130 (2012).
13. Ghaemi, P. & Wilczek, F. *Physica Scripta* **T146,** 014019 (2012).
14. Sau, J. D., Lutchyn, R. M., Tewari, S. & Das Sarma, S. *Phys. Rev. Lett.* **104,** 040502 (2010).
15. Oreg, Y., Refael, G. & von Oppen, F. *Phys. Rev. Lett.* **105,** 177002 (2010).
16. Nayak, C. & Wilczek, F. *Nucl. Phys. B* **479,** 529–553 (1996).
17. Moore, G. & Read, N. *Nucl. Phys. B* **360,** 362–396 (1991).
18. Ivanov, D. A. *Phys. Rev. Lett.* **86,** 268–271 (2001).

MICROSCOPY

Reconstructing the third dimension

An approach to microscopy has been developed that can be used to determine, from a single imaging angle, both the position of a specimen's individual atoms in the plane of observation and the atoms' vertical position. SEE LETTER P.243

DILANO SALDIN

In this issue (page 243)[1], Van Dyck and Chen report an electron-microscopy technique that draws on an analogy with the method used in astronomy for determining distances to galaxies. The technique allows high-resolution, three-dimensional information to be obtained about a sample using only one viewing direction.

A central observation in astronomy is that distant galaxies are moving away from us, and from each other, with a speed that is proportional to their distance from Earth. In other words, the farther away they are, the faster they are moving. Because the speeds of galaxies can be measured from the Doppler effect, which shifts the galaxies' light to the red end of the electromagnetic spectrum, their distances can be determined using the constant of proportionality between speed and distance, known as the Hubble constant[2]. This central observation, called Hubble's law, is crucial evidence for the now accepted view that the Universe originated in a Big Bang, as a tiny, unimaginably dense entity that has been expanding ever since. What's more, it provides astronomers with a neat way of determining the distances to objects, which would otherwise be impossible using only the objects' observed positions in the sky.

The problem of calculating the distance of an object from the plane of observation (the image plane) also exists in other fields of research. One such field is electron microscopy. Because an electron microscope produces a two-dimensional image of a sample, the lateral positions of the sample's atoms can be observed directly to high precision. However, as with most images recorded on two-dimensional media, including conventional cameras, the vertical positions (heights) of the specimen's constituents in the direction perpendicular to the image plane are not easily found.

One way to determine these heights involves tomographic electron microscopy, in which a three-dimensional image is reconstructed from projected two-dimensional images obtained from a range of viewing directions. But for imaging at atomic resolution, this method would require the sample to be mechanically tilted with sub-ångström precision over a large angular range, a feat that has not yet been achieved. Moreover, a typical electron microscope records only the intensity of the electron wave that is scattered by the sample. The wave's phase provides much

of the important structural information about the sample, but the phase at different image points is not ordinarily recorded on an electron micrograph. However, it can be found using a process known as focal series reconstruction[3,4], or by creating interference between the scattered wave and a known (unscattered) reference wave, as is done in an imaging technique called electron holography[5].

Van Dyck and Chen suggest that it is possible to calculate the heights of atoms from the image plane using only one viewing direction, if this phase can also be determined. The authors point out that the phase speed — the rate at which the phase changes with scattering angle — near the image of a particular atom is approximately proportional to the height of that atom. This approximate proportionality then allows the height to be determined, in rough analogy to Hubble's law. However, it should be noted that the proportionality is strictly valid only in the vicinity of the atom's projected position in the image plane, and that its value is approximate because of aberrations in the microscope's lenses. Therefore, the local and approximate constant of proportionality allows only the height of that particular atom to be measured.

In applications of holography that have captured the public's imagination, a ghostly three-dimensional image of a macroscopic object is reconstructed from the information given by a two-dimensional interference pattern (the hologram) of two laser beams. In this form of imaging, which was first proposed[6] by the physicist Dennis Gabor to image microscopic objects using electron waves instead of light beams, the macroscopic object may be regarded as an ensemble of point scatterers, and the fact that fringes in the interference pattern overlap is no obstacle to correctly reconstructing the spatial positions of these point scatterers. In this technique, the three-dimensional nature of the image automatically gives information about the third dimension. This is also the basis of atomic-source holography[7], which has been used, for example, to determine the positions (including the heights) of atoms adsorbed on a surface.

Van Dyck and Chen propose that knowledge about the third dimension can instead be obtained by first reconstructing the normally invisible phase of the electron wave, and then exploiting the analogy with Hubble's law. They demonstrate their technique for atoms in a system composed of two layers of graphene — a one-atom-thick, honeycomb-like lattice of carbon. But for the proposed technique to be more generally applicable, it will be necessary to extend the algorithm to reconstruct truly three-dimensional objects, as Gabor envisaged. ■

Dilano Saldin *is in the Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53211, USA.*
e-mail: dksaldin@uwm.edu

1. Van Dyck, D. & Chen, F.-R. *Nature* **486,** 243–246 (2012).
2. Hubble, E. *Proc. Natl Acad. Sci. USA* **15,** 168–173 (1929).
3. Coene, W. M. J., Thust, A., Op de Beeck, M. & Van Dyck, D. *Ultramicroscopy* **64,** 109–135 (1996).
4. Hsieh, W.-K., Chen, F.-R., Kai, J.-J. & Kirkland, A. I. *Ultramicroscopy* **98,** 99–114 (2004).
5. Lehmann, M. & Lichte, H. *Microsc. Microanal.* **8,** 447–466 (2002).
6. Gabor, D. *Nature* **161,** 777–778 (1948).
7. Saldin, D. K. *Surf. Rev. Lett.* **4,** 441–457 (1997).

PLANT IMMUNOLOGY

A life or death switch

The identification of two receptors for salicylic acid reveals how the hormone controls cell death and survival during plant immune responses, in tissues close to and distant from the site of infection. See Letter p.228

ANDREA A. GUST & THORSTEN NÜRNBERGER

Immunity to microbial infection is an inherent feature of multicellular organisms. In plants, immune responses are activated when cellular receptors recognize microbial proteins (effectors) that betray the invader to the plant's surveillance system[1,2]. This activation requires the plant hormone salicylic acid, which is produced on microbial attack[3]. But how plants detect the hormone, and how it performs its immunity-associated functions, has remained unclear. On page 228 of this issue, Fu *et al.*[4] report the identification of two salicylic acid receptors in the model plant *Arabidopsis thaliana*, and provide a fascinating explanation of how the hormone controls both cell death at the site of infection, and cell survival and immune activation in non-infected tissues.

In plants, effector-triggered immunity (ETI) is often accompanied by programmed cell death (PCD) at the infection site. In addition to participating in local immune responses, PCD triggers long-lasting immunity against a broad spectrum of microbes throughout the plant — a protective mechanism referred to as systemic acquired resistance[3]. Salicylic acid participates in these immune responses by controlling the movement of a protein called NPR1 (non-expresser of pathogenesis-related genes 1) from the cell cytoplasm to the nucleus[5]. Once in the nucleus, NPR1 regulates the expression of plant defence genes. Because mutant

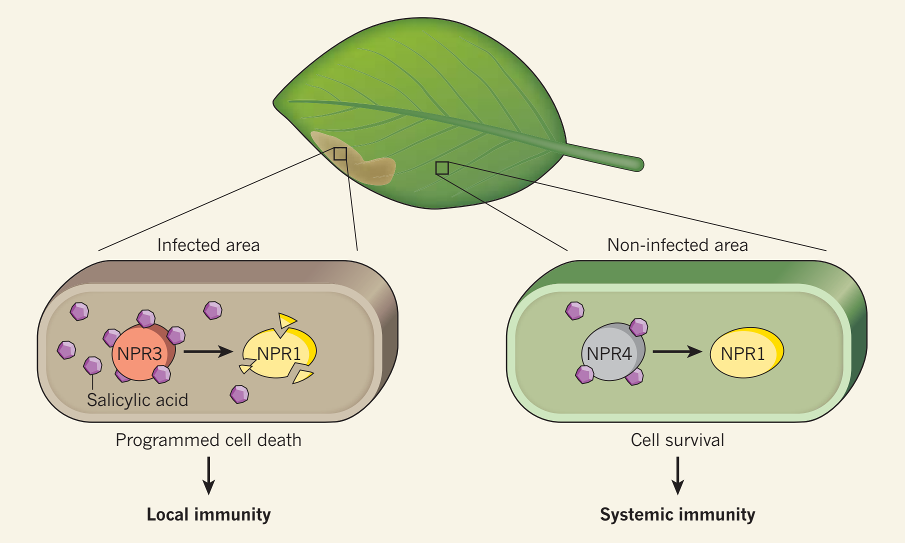


Figure 1 | Salicylic acid-mediated control of plant cell death and survival. On microbial infection, levels of the plant hormone salicylic acid increase, with its concentration decreasing gradually with increasing distance from the site of infection. Fu *et al.*[4] show that at high salicylic acid concentrations — typically found in infected areas of the plant — the receptor NPR3, which binds salicylic acid with low affinity, mediates degradation of the cell-death suppressor NPR1 (left panel), thereby favouring programmed cell death and local effector-triggered immunity. However, at the lower salicylic acid concentrations typically found in cells distant from the infection site, salicylic acid cannot bind to the low-affinity receptor NPR3, so cell death is blocked. In these cells, salicylic acid instead binds to the high-affinity receptor NPR4 (right panel), blocking degradation of NPR1, and thereby favouring cell survival and the expression of genes associated with systemic immunity.

plants that are insensitive to salicylic acid and plants that lack NPR1 exhibit similar immune defects, NPR1 was previously proposed[6] to be a salicylic acid receptor. However, Fu *et al.*[4] did not detect any physical interaction between salicylic acid and NPR1, suggesting that NPR1 does not serve this receptor function.

What, then, could be the bona fide salicylic acid receptor mediating local and systemic immune activation in plants? The research group presenting the current paper has previously shown[7] that the proper functioning of NPR1 requires that the protein is broken down by cellular protein-degradation machinery called the proteasome. So Fu and colleagues hypothesized that adaptor proteins that link NPR1 to the proteasome might be receptors for salicylic acid. Two members of the NPR protein family, NPR3 and NPR4, exhibit a protein-domain structure that is characteristic of such adaptor proteins, leading the authors to surmise that these proteins could be the proteasome adaptors that mediate NPR1 degradation. To validate this assumption, the researchers demonstrated that NPR1 is degraded by the proteasome in wild-type *A. thaliana* plants, but not in plants in which the genes for NPR3 and NPR4 have both been knocked out.

Fu and colleagues used *in vitro* protein–protein interaction studies to assess the effect of salicylic acid on the formation of protein complexes between NPR1 and NPR3 or NPR4. The authors found, surprisingly, that salicylic acid promotes NPR1–NPR3 interaction, but disrupts formation of the NPR1–NPR4 complex. Thus it seems that salicylic acid interacts physically with NPR3 and NPR4 in a receptor-like manner, but that this interaction has opposing effects on the adaptor proteins' interactions with NPR1. The authors also found that although NPR3 and NPR4 both bind to salicylic acid, NPR4 binds with greater affinity than does NPR3. Hence, *Arabidopsis* plants contain two salicylic acid receptors, NPR3 and NPR4, which differ in their affinity for the hormone and in their roles in NPR1 degradation, with NPR3 mediating NPR1 breakdown only in the presence of salicylic acid and NPR4 only in its absence.

What are the biological consequences of NPR3- or NPR4-mediated degradation of NPR1? Fu and colleagues found that both local PCD and local ETI responses to bacterial infection were compromised in plants lacking the genes encoding both NPR3 and NPR4. The impairment of PCD, combined with the fact that NPR1 accumulates in the mutant plants (because it cannot be degraded) suggests that NPR1 suppresses PCD in wild-type plants. Because salicylic acid levels are highest at infection sites[8], Fu *et al.* propose that binding of the hormone to the lower-affinity receptor NPR3 mediates NPR1 degradation and de-repression of PCD and ETI in infected cells (Fig. 1).

However, infection causes salicylic acid levels to increase systemically as well as locally, with its concentration decreasing gradually with increasing distance from the infection site[8]. In cells farther away from the infected area, salicylic acid levels are likely to drop below the concentration required for NPR3-mediated NPR1 degradation and, thus, PCD. Fu and colleagues propose that, in these cells, salicylic acid binds instead to the higher-affinity receptor NPR4, which inhibits NPR4-mediated NPR1 degradation and thereby facilitates NPR1 accumulation, cell survival and subsequent salicylic acid-dependent gene expression (Fig. 1). Consistent with this model, the authors showed that NPR1 levels are lowest in cells undergoing PCD and highest in cells surrounding PCD lesions.

Several mutant plants that exhibit runaway PCD have been identified[9], and the question of how plants control PCD has been a major area of research. Fu and colleagues' findings provide compelling evidence that salicylic acid acts as an immune signal to determine cell fate in plant immunity. Studies investigating plant proteins associated with abnormal PCD should now examine whether these proteins might contribute to the functionality of NPR3 or NPR4.

Salicylic acid is the only major plant hormone for which the receptor has remained elusive. Fu and colleagues' demonstration that the two salicylic acid receptors control distinct defence strategies by de-repressing local cell death and immunity at the infection site in one case, and systemic immunity remote from the infection site in the other, is reminiscent of the de-repression of physiological programs enacted by other plant hormones, such as auxin, gibberellic acid and jasmonic acid[10]. However, NPR3 and NPR4 are the first plant-hormone sensors for which differences in binding affinity have been shown to mediate differential control of plant responses. Because most plant hormones regulate multiple aspects of plant life, it is certainly possible that other plant hormone receptors use a comparable mode of action. Consistent with this idea is the recent identification[11] of auxin hormone-binding proteins that have different ligand affinities, suggesting that plants also have means for differential sensing of auxin[1]. ■

Andrea A. Gust *and* **Thorsten Nürnberger** *are in the Centre of Plant Molecular Biology, Department of Plant Biochemistry, University of Tübingen, D-72076 Tübingen, Germany.*
e-mails: nuernberger@uni-tuebingen.de; andrea.gust@zmbp.uni-tuebingen.de

1. Dodds, P. N. & Rathjen, J. P. *Nature Rev. Genet.* **11,** 539–548 (2010).
2. Jones, J. D. & Dangl, J. L. *Nature* **444,** 323–329 (2006).
3. Spoel, S. H. & Dong, X. *Nature Rev. Immunol.* **12,** 89–100 (2012).
4. Fu, Z. Q. *et al. Nature* **486,** 228–232 (2012).
5. Mou, Z., Fan, W. & Dong, X. *Cell* **113,** 935–944 (2003).
6. Cao, H., Glazebrook, J., Clarke, J. D., Volko, S. & Dong, X. *Cell* **88,** 57–63 (1997).
7. Spoel, S. H. *et al. Cell* **137,** 860–872 (2009).
8. Enyedi, A. J., Yalpani, N., Silverman, P. & Raskin, I. *Proc. Natl Acad. Sci. USA* **89,** 2480–2484 (1992).
9. Lam, E. *Nature Rev. Mol. Cell Biol.* **5,** 305–315 (2004).
10. Robert-Seilaniantz, A., Grant, M. & Jones, J. D. *Annu. Rev. Phytopathol.* **49,** 317–343 (2011).
11. Calderón Villalobos, L. I. *et al. Nature Chem. Biol.* **8,** 477–485 (2012).

ASTRONOMY

An infant giant

Spectroscopic measurements of a galaxy that shines brightly at submillimetre wavelengths place it in the middle of a nascent galaxy cluster at a scant one billion years after the Big Bang. **SEE LETTER P.233**

ALBERTO D. BOLATTO

As in many scientific fields, but perhaps more than in most, advances in technology frequently drive progress in astronomy. The deployment of a new instrument or a capability can afford a completely different view of the Universe, opening a window onto an aspect of reality that was previously unsuspected or only theorized. One such advance occurred in the late 1990s with the advent of large-format submillimetre-wave cameras — particularly SCUBA, the Submillimetre Common-User Bolometer Array mounted on the James Clerk Maxwell Telescope in Mauna Kea, Hawaii — and the resulting discovery of a class of luminous yet elusive galaxies. On page 233 of this issue, Walter *et al.*[1] describe an analysis that advances our understanding of this family of galaxies and closes a chapter on the story of their origins.

Early on, images taken with SCUBA revealed a population of galaxies that shine brightly at submillimetre wavelengths, prosaically named submillimetre galaxies (SMGs). One of the first deep SCUBA images to uncover SMGs was obtained[2] in 1998 through observations of the Hubble Deep Field (Fig. 1), which is perhaps the most emblematic patch of sky observed by the Hubble Space Telescope. Surprisingly, SMGs were very faint, or

even invisible, in optical images of this and other fields. Moreover, the SCUBA data provided no spectroscopic information, and thus no direct knowledge of the objects' distance or redshift, which are necessary to establish most of a galaxy's properties. What were these mysterious submillimetre sources?

The poster child for this question is HDF 850.1, the brightest SMG in the Hubble Deep Field. Walter and colleagues determine the galaxy's redshift, providing an important piece of the SMG puzzle. Analyses of the galaxies present in optical images at positions coincident, within errors, with those of SCUBA sources suggested from the outset that these sources are located at cosmological distances from Earth, corresponding to a time when the Universe was a fraction of its present age. The analyses also indicated that SMGs are tremendously powerful systems — with energy outputs of the order of 10^{12} times that of the Sun[2,3]. Such luminosities imply that, if they are powered mostly by star formation, SMGs are forming stars at rates several hundred times that of our Milky Way. Yet dust hides these starbursts from optical observations, which could therefore be missing an important fraction of the star-formation activity occurring in the early Universe[3].

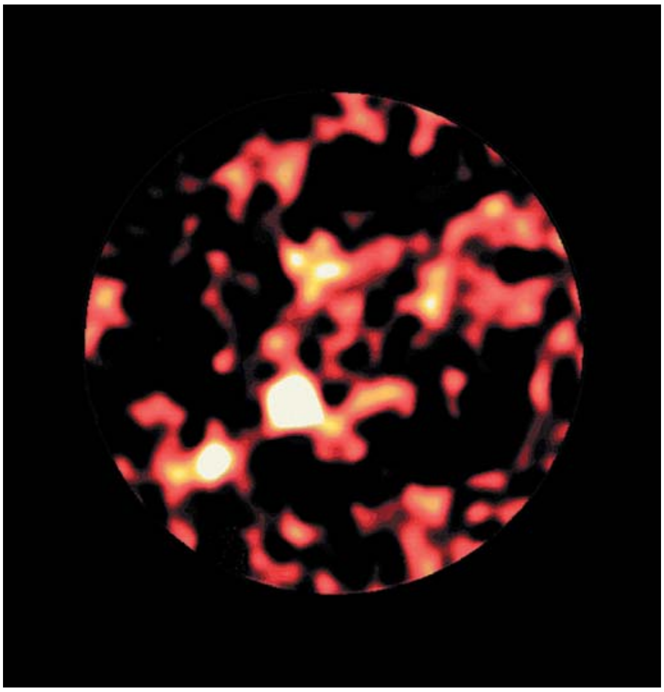


Figure 1 | Structure in the infant Universe. The brightest source in this submillimetre-wavelength image of the Hubble Deep Field, obtained using the SCUBA camera[2], is a galaxy called HDF 850.1. Walter and colleagues' spectroscopic observations[1] indicate that this object is at the centre of a nascent galaxy cluster at redshift 5.183. White and yellow denote bright sources; red represents fainter sources. The circle's diameter is about 200 arcseconds (one-tenth of a full Moon).

REF. 2

Although there are indirect indications of how far away SMGs are located, spectroscopy of these sources — and so direct measurement of their redshifts — has remained difficult. One successful indirect technique relies on the correlation between an object's radio and far-infrared luminosities, and uses radio interferometers to pinpoint its precise sky coordinates. Measurement of the source's redshift then requires long-exposure spectroscopy of its optical counterparts using the largest optical telescopes. The results from studies based on this indirect method indicate that most SMGs exist at redshifts of about 1.5–3, around the peak of star-formation activity in the Universe[4]. Because of the built-in observational selection effects of this technique, the question of how many SMGs lie at redshifts beyond 3 is still open. SMGs could shine in the submillimetre range at much greater distances than in the radio range, and so this approach might fail to detect distant SMGs.

In their study, Walter and collaborators used the IRAM Plateau de Bure Interferometer near Montmaur, France, to obtain direct spectroscopic measurements of HDF 850.1, pinpointing its redshift at 5.183 — a scant 1.1 billion years after the Big Bang. The technique relies on detecting the spectral lines associated with two rotational transitions of the carbon monoxide (CO) molecule to unambiguously determine the redshift. The method has recently become feasible thanks to the increasing information-transmission speed and computing power of (sub)millimetre-wave interferometric spectrographs, which can provide instantaneous spectra spanning tens of gigahertz. Determination of redshifts at millimetre and submillimetre wavelengths is a powerful tool, and will be used increasingly for the study of the early Universe with new advanced observatories, particularly the Atacama Large Millimetre Array in Chile's Atacama Desert.

The authors' spectroscopic analysis also revealed a spectral line associated with ionized carbon. This detection not only confirms the SMG redshift, but also supports the idea that the radiation previously detected by SCUBA stems from a monstrous starburst producing stars 800 times faster than the present-day Milky Way. Yet the source is completely invisible in the deepest optical images, in which the radiation from young stars is not detected because of obscuration by dust present in the source. This obscuring dust absorbs ultraviolet and optical light and re-emits it at far-infrared wavelengths, which are subsequently 'redshifted' (stretched) by the expansion of the Universe and so fall in the submillimetre band observed by SCUBA.

The authors find that the abundance of dust relative to gas in HDF 850.1 is similar to that found in the Milky Way. Dust is predominantly composed of heavy elements (elements other than hydrogen and helium), most of which originate from stellar explosions known as supernovae, which are caused by the death of massive stars. The unexpectedly high abundance of dust relative to gas at this early cosmic epoch indicates that heavy-element enrichment of the interstellar medium of HDF 850.1 proceeded at an extremely fast pace. The mechanism responsible for creating this much dust so quickly is not entirely understood.

A remarkable finding is that HDF 850.1 is not alone. It is sitting in the middle of a clear 'overdensity' of sources that also includes a rare quasar (an active galactic nucleus powered by a supermassive black hole). Twelve other sources have been identified within a redshift interval of about 0.03 from the galaxy's location. This interval corresponds to a local depth in the sky of about 2.5 megaparsecs (8 million light years), which is also the approximate extent of the structure in the sky. This overdensity is one of two[5] spectroscopically verified candidate protoclusters at such high redshift, and thus represents one of the earliest structures known in the Universe — occurring after the epoch at which relic radiation from the Big Bang, known as the cosmic microwave background radiation, emerged. This discovery lends further support to the idea that the most luminous SMGs are signposts for rare overdensities in the overall structure of the Universe. Such extremely overdense regions are thought to have evolved into present-day massive clusters, and the brightest SMGs in these regions probably became today's giant elliptical galaxies at the centres of the clusters' potential well[6]. The conclusion of the story for HDF 850.1, then, is that this SMG is nothing less than an infant giant undergoing a growth spurt. ■

Alberto D. Bolatto *is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.*
e-mail: bolatto@astro.umd.edu

1. Walter, F. *et al. Nature* **486,** 233–236 (2012).
2. Hughes, D. H. *et al. Nature* **394,** 241–247 (1998).
3. Blain, A. W., Smail, I., Ivison, R. J. & Kneib, J.-P. *Mon. Not. R. Astron. Soc.* **302,** 632–648 (1999).
4. Chapman, S. C., Blain, A. W., Smail, I. & Ivison, R. J. *Astrophys. J.* **622,** 772–796 (2005).
5. Capak, P. L. *et al. Nature* **470,** 233–235 (2011).
6. Tacconi, L. J. *et al. Astrophys. J.* **680,** 246–262 (2008).

ARTICLE

doi:10.1038/nature11204

Visualizing heavy fermions emerging in a quantum critical Kondo lattice

Pegor Aynajian[1]*, Eduardo H. da Silva Neto[1]*, András Gyenis[1], Ryan E. Baumbach[2], J. D. Thompson[2], Zachary Fisk[3], Eric D. Bauer[2] & Ali Yazdani[1]

In solids containing elements with *f* orbitals, the interaction between *f*-electron spins and those of itinerant electrons leads to the development of low-energy fermionic excitations with a heavy effective mass. These excitations are fundamental to the appearance of unconventional superconductivity and non-Fermi-liquid behaviour observed in actinide- and lanthanide-based compounds. Here we use spectroscopic mapping with the scanning tunnelling microscope to detect the emergence of heavy excitations with lowering of temperature in a prototypical family of cerium-based heavy-fermion compounds. We demonstrate the sensitivity of the tunnelling process to the composite nature of these heavy quasiparticles, which arises from quantum entanglement of itinerant conduction and *f* electrons. Scattering and interference of the composite quasiparticles is used to resolve their energy–momentum structure and to extract their mass enhancement, which develops with decreasing temperature. The lifetime of the emergent heavy quasiparticles reveals signatures of enhanced scattering and their spectral lineshape shows evidence of energy–temperature scaling. These findings demonstrate that proximity to a quantum critical point results in critical damping of the emergent heavy excitation of our Kondo lattice system.

A local magnetic moment occurs when a strongly interacting quantum state, such as an atomic *d* or *f* orbital, cannot be doubly occupied owing to strong on-site Coulomb repulsion[1]. In the presence of a dilute concentration of such magnetic moments in a metal, spin-flip scattering of conduction electrons from these local moments results in their collective magnetic screening below a characteristic temperature called the Kondo temperature, T_K (Fig. 1a)[2]. In materials where local moments are arranged in a dense periodic array, the so-called Kondo lattice, the deconfinement of localized orbitals through their hybridization with the conduction electrons results in composite low-energy excitations with a heavy effective mass (Fig. 1b). Tuning the hybridization between *f* orbitals and itinerant electrons can destabilize the heavy Fermi-liquid state towards an antiferromagnetically ordered ground state at a quantum critical point (QCP)[3–8]. In proximity to such a quantum phase transition—between itinerancy and localization of *f* electrons—many heavy-fermion systems exhibit unconventional superconductivity at low temperatures (Fig. 1c)[9].

Thermodynamic and transport studies have long provided evidence for heavy quasiparticles, their unconventional superconductivity and non-Fermi-liquid behaviour in a variety of material systems[9–14]. However, the emergence of a coherent band of heavy quasiparticles near the Fermi energy in a Kondo lattice system is still not well understood[14–17]. Part of the challenge has been the inability of spectroscopic measurements to probe the development of heavy quasiparticles with lowering of temperature and to characterize their properties with high energy resolution. Such precise measurements of heavy-fermion formation are not only required for understanding the nature of these electronic excitations close to quantum phase transitions[18], but are also critical to identifying the source of unconventional superconductivity near such transitions.

Composite heavy-fermion excitations

The emergence of composite heavy fermions in a Kondo lattice can be considered as a result of the hybridization of two electronic bands: one

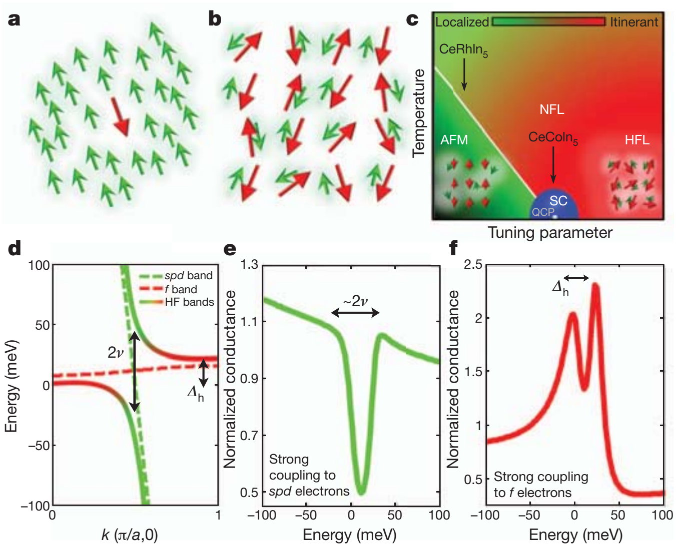


Figure 1 | Tunnelling into a Kondo lattice. a, **b**, Schematic representations of a single-impurity Kondo effect (**a**) and a Kondo lattice (**b**), illustrating the screening (hybridization) of the local moments (red arrows) by the itinerant conduction electrons (green arrows). **c**, Schematic phase diagram of heavy-fermion systems where the electronic ground state can be tuned from antiferromagnetism (AFM) with localized *f* moments to a heavy Fermi liquid (HFL) with itinerant *f* electrons. At low temperatures, superconductivity (SC) sets in near the quantum critical point (QCP) from a non-Fermi liquid (NFL). Insets show cartoon pictures of the spin fluctuations in the different phases. **d**, Bare electronic bands (dashed lines; *spd* and *f*) and hybridized heavy-fermion bands (HF; solid lines) displaying a direct (2ν) and an indirect (Δ_h) hybridization gap. **e**, Tunnelling spectrum computed from the hybridized band structure in **d** for a tunnelling ratio $t_f/t_c = -0.025$, showing sensitivity to the direct hybridization gap (2ν). **f**, Similar spectra computed with $t_f/t_c = -0.37$, showing sensitivity to the indirect gap (Δ_h). See Supplementary Information section I for details of the model.

[1]Joseph Henry Laboratories and Department of Physics, Princeton University, Princeton, New Jersey 08544, USA. [2]Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. [3]Department of Physics and Astronomy, University of California, Irvine, California 92697, USA.
*These authors contributed equally to this work.

dispersing band due to conduction electrons and one weakly dispersing band originating from localized f electrons (dashed lines in Fig. 1d). This hybridization generates low-energy quasiparticles that are a mixture of conduction electrons and f electrons with a modified band structure characterized by the so-called direct ($2v$) and indirect (Δ_h) hybridization gaps, as shown in Fig. 1d[17,19]. Various theoretical approaches, including several numerical studies, reproduce the generic composite band structure shown in Fig. 1d[20–24]. Recent theoretical modelling has also shown that tunnelling spectroscopy can be a powerful probe of this composite nature of heavy fermions[25–28]. Depending on the relative tunnelling amplitudes to the light conduction (t_c) or to the heavy f-like (t_f) components of the composite quasiparticles, and due to their interference, tunnelling spectroscopy can be sensitive to different features of the hybridized band structure. Figure 1d–f shows examples of model calculations (see Supplementary Information section I) illustrating the sensitivity of the spectra to predominant tunnelling to the light (Fig. 1e) or heavy (Fig. 1f) electronic states.

Recent advances in the application of scanning tunnelling microscopy (STM) to heavy fermions are providing a new approach to examining the correlated electrons in these systems with high energy and spatial resolutions. STM and point-contact experiments on heavy-fermion compounds have shown evidence for hybridization of the conduction electrons with the f orbitals and have been used to probe the so-called hidden order phase transition involving heavy f electrons in URu_2Si_2 (refs 29–32). Sudden onset of the hidden order phase seems to give rise to strong modification of the band structure in URu_2Si_2 as detected by STM measurements[30,31]. However, these changes are correlated with the phase transition into the hidden order at 17.5 K rather than being the generic physics of heavy Fermi liquids that should appear at higher temperatures and evolve smoothly with lowering of temperature. Direct experimental observation of the gradual formation of heavy quasiparticles with decreasing temperature and evidence of their composite nature, which is ubiquitous to all heavy fermions, as well as examination of their properties in proximity to QCPs, have remained out of the reach of STM and other spectroscopic measurements.

$CeMIn_5$ as a model heavy-fermion system

To provide a controlled study of the emergence of heavy-fermion excitations within a Kondo lattice system that can be tuned close to a QCP, we carried out studies on the $CeMIn_5$ (with M = Co, Rh) material system. These so-called 115 compounds (the chemical formula could be written $Ce_1M_1In_5$) offer the possibility of tuning the interaction between the f orbitals of Ce and the itinerant *spd* conduction electrons using isovalent substitutions at the transition metal site within the same tetragonal crystal structure. Consequently, the ground state of this system can be tuned (in stoichiometric compounds) between antiferromagnetism, as in $CeRhIn_5$ (Néel temperature $T_N = 3.5$ K), to superconductivity, as observed in $CeCoIn_5$ (superconducting transition temperature $T_c = 2.3$ K) and $CeIrIn_5$ ($T_c = 0.4$ K)[9]. Previous studies indicate that $CeCoIn_5$ is very close to a QCP[33–36], whereas $CeRhIn_5$ can be tuned close to this transition with application of pressure[7,37]. These experiments confirm that superconductivity in the 115 system emerges at low temperatures close to a QCP from heavy low-energy excitations that developed at high temperature[7,9,36]. More specifically, transport studies show a drop in the electrical resistivity of $CeCoIn_5$ around 50 K (which has been interpreted as evidence for the development of a coherent heavy quasiparticle band) followed by a T-linear resistivity at lower temperature (above T_c)[38] — a behaviour that has been associated with the proximity to the QCP. Quantum oscillations and thermodynamic measurements find a heavy electron effective mass (10–50 m_0, where m_0 is the bare electron mass) for $CeCoIn_5$, whereas in the same temperature range the f electrons in $CeRhIn_5$ are effectively decoupled from the conduction electrons[39,40].

Figure 2 shows STM images of a single crystal of $CeCoIn_5$ that has been cleaved *in situ* in our variable temperature ultrahigh-vacuum STM. In this family of compounds, the cleaving process results in the exposure of multiple surfaces terminated with different chemical compositions. The crystal symmetry necessarily requires multiple surfaces for cleaved samples, as no two equivalent consecutive layers occur within the unit cell. Therefore breaking of any single chemical bond will result in different layer terminations on the two sides of the cleaved sample. Experiments on many cleaved samples have revealed three different surfaces, two of which are atomically ordered (termed surfaces A and B in Fig. 2a, b) with a periodicity of ~4.6 Å corresponding to the lattice constant of the bulk crystal structure, whereas the third surface (termed surface C, Fig. 2b) is reconstructed. Comparison of the relative heights of the sub-unit-cell steps between the different layers (Fig. 2c, d) to the crystal structure determined from scattering experiments[41] suggests that exposed surfaces A, B and C correspond to the Ce–In, Co and In_2 layers, respectively. Experiments on $CeCo(In_{0.9985}Hg_{0.0015})_5$ and $CeRhIn_5$ reveal similar results, where cleaving exposes the corresponding multiple layers in those compounds (see Supplementary Information section II). Hg defects in $CeCoIn_5$ at this concentration have negligible influence on its thermodynamic and transport properties and are introduced for the scattering experiments described below[42].

Signatures of hybridization and composite excitations

Spectroscopic measurements of $CeCoIn_5$ show the sensitivity of the tunnelling process to the composite nature of the hybridized heavy-fermion states. Tunnelling spectra on surface A (identified as the Ce–In layer) of $CeCo(In_{0.9985}Hg_{0.0015})_5$ show that on cooling the sample dramatic changes develop in the spectra in an asymmetric fashion about the Fermi energy (Fig. 3a). (The same behaviour is also observed in $CeCoIn_5$; see Supplementary Information section III.)

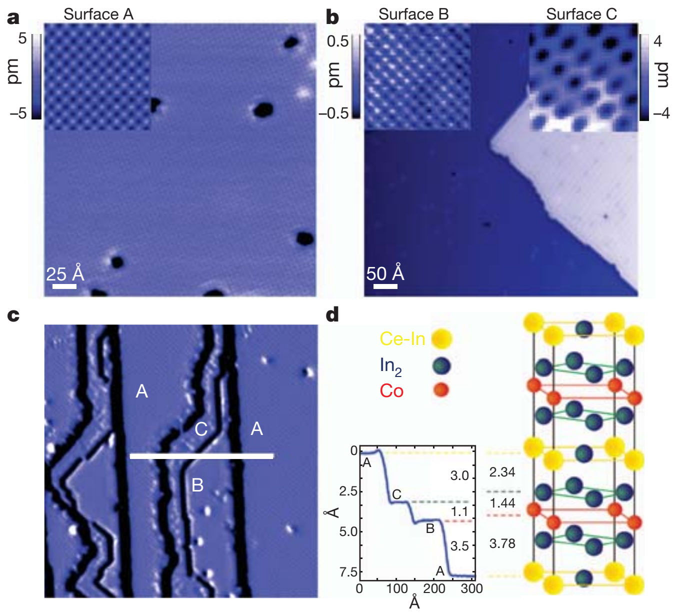


Figure 2 | STM topographies on $CeCoIn_5$. **a**, Constant current topographic image (+200 mV, 200 pA) showing an atomically ordered surface (termed surface A) with a lattice constant of ~4.6 Å. **b**, Topographic image (−200 mV, 200 pA) showing two consecutive layers: a distinct atomically ordered surface (termed surface B, lattice constant ~4.6 Å, dark blue) and a reconstructed surface (termed surface C, light blue). Insets in **a** and **b** show magnified images (45 × 45 Å^2) of the three different surfaces. **c**, Constant current topographic image (−150 mV, 365 pA) displaying all three surfaces (the derivative of the topography is shown to enhance contrast). **d**, A line section through the different surfaces (solid line in **c**) showing the relative step heights (left) compared to the bulk crystal structure (right).

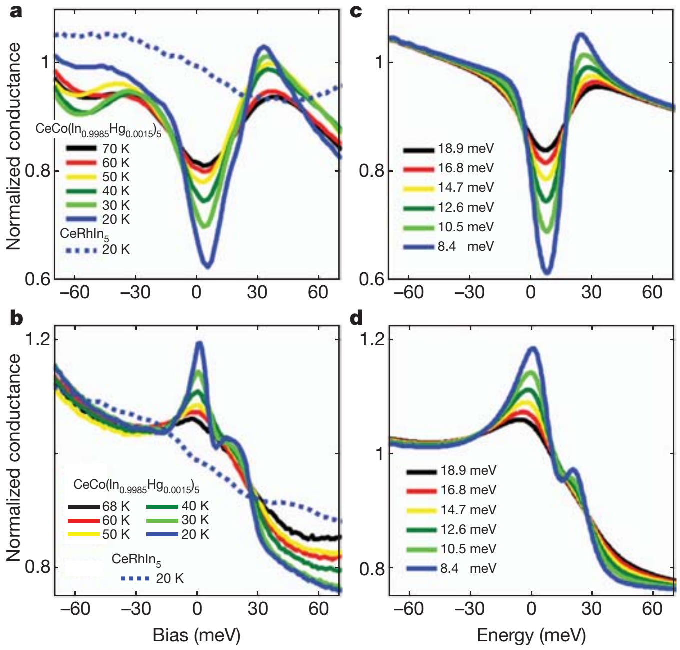


Figure 3 | Composite nature of heavy-fermion excitations. **a** Averaged tunnelling spectra (−150 mV, 200 pA) measured on surface A of $CeCo(In_{0.9985}Hg_{0.0015})_5$ for different temperatures (T, in K; solid lines) and on the corresponding surface A of $CeRhIn_5$ at 20 K (dashed line). **b**, Averaged tunnelling spectra (−150 mV, 200 pA) measured on surface B of $CeCo(In_{0.9985}Hg_{0.0015})_5$ for different temperatures (T, in K; solid lines) and on corresponding surface B of $CeRhIn_5$ at 20 K (dashed line). **c**, **d**, Tunnelling spectra computed for $t_f/t_c = -0.01$ (**c**) and $t_f/t_c = -0.20$ (**d**) for selected values of γ_f (in meV; solid lines). See Supplementary Information section I for details of the model.

The redistribution of the spectra observed on this surface is consistent with a tunnelling process that is dominated by coupling to the light conduction electrons and displays signatures of the direct hybridization gap ($2v$) experienced by this component of the heavy-fermion excitations (for example, see Fig. 1d, e). In contrast to these observations, similar measurements on the corresponding surface of $CeRhIn_5$ show spectra that are featureless in the same temperature range (Fig. 3a, dashed line) and are consistent with the more localized nature of the Ce f orbitals in $CeRhIn_5$ as compared to $CeCoIn_5$. The hybridization gap structure in $CeCoIn_5$ is also centred above the chemical potential (8 meV, see Fig. 3a), which makes access difficult for angle-resolved photoemission experiments[43–45]—the typical technique used for probing electronic band structure in solids.

The composite nature of the heavy-fermion excitations manifests itself by displaying different spectroscopic characteristics for tunnelling into the different atomic layers. Figure 3b shows spectra measured on surface B (identified as Co) of $CeCo(In_{0.9985}Hg_{0.0015})_5$ that look very different from those measured on surface A (Fig. 3a). In the temperature range where spectra on surface A (Fig. 3a) develop a depletion of spectral weight near the Fermi energy, surface B shows a sharp enhancement of spectral weight within the same energy window (Fig. 3b). With further lowering of temperature, the enhanced tunnelling on surface B evolves into a double-peak structure. As a control experiment, measurements on the corresponding surface in $CeRhIn_5$, once again, display no sharp features in the same temperature and energy windows (Fig. 3b, dashed line). The spectroscopic features of surface B of $CeCo(In_{0.9985}Hg_{0.0015})_5$ display the characteristic signatures of dominant tunnelling to the f component of the heavy quasiparticles, which reside near the Fermi energy and are expected to display the indirect hybridization gap (Δ_h; see Fig. 1d, f).

Modelling the tunnelling to composite heavy excitations can reproduce our spectroscopic measurements on the two different atomically ordered surfaces of $CeCo(In_{0.9985}Hg_{0.0015})_5$. Following recent theoretical efforts[26,27], we compute spectroscopic properties of a model band structure in which a single hole-like itinerant band of spd-like electrons hybridizes with a narrow band of f-like electrons (see Supplementary Information section I for details of the model). The results of our calculations (Fig. 3c, d) are sensitive to the ratio of tunnelling (t_f/t_c) into the heavy f states to tunnelling into the light conduction band—a behaviour that explains the differences between the tunnelling processes on the different cleaved surfaces (Fig. 3a, b). Although naively one would expect that tunnelling to the heavy excitations would be more pronounced on the Ce–In layer, recent first principles calculations show that the amplitude of the hybridization of the f states with the out-of-plane spd electrons can be remarkably larger than the amplitude of the hybridization with the in-plane spd electrons[21].

Visualizing quasiparticle mass enhancement

To directly probe the energy–momentum structure of heavy quasiparticles in the 115 material systems, we have carried out spectroscopic mapping with the STM that enables us to visualize the scattering and interference of these quasiparticle excitations from impurities or structural defects. Elastic scattering of quasiparticles from these imperfections gives rise to standing waves in the conductance maps at wavelengths corresponding to $2\pi/q$, where $\mathbf{q} = \mathbf{k}_f - \mathbf{k}_i$ is the momentum transfer between initial ($\mathbf{k}_i$) and final ($\mathbf{k}_f$) states at the same energy. We expect that those $\mathbf{q}$ with the strongest intensity connect regions of high density of states on the contours of constant energy, and hence provide energy–momentum information about the quasiparticle excitations. We characterize the scattering $\mathbf{q}$ using discrete Fourier transforms (DFTs) of STM conductance maps measured at different energies. The presence of Hg substitutions in $CeCo(In_{0.9985}Hg_{0.0015})_5$ provides a sufficient number of scattering centres to enhance signal to noise ratio for such quasiparticle interference (QPI) measurements.

Figure 4a shows examples of energy-resolved STM conductance maps on surface A of $CeCo(In_{0.9985}Hg_{0.0015})_5$ measured at 20 K; the maps display signatures of scattering and interference of quasiparticles from defects and step edges. These conductance maps show clear changes of the wavelength of the modulations as a function of energy. Perhaps the most noticeable are the changes around each random defect (see Supplementary Information section II for the corresponding STM image showing the location of the Hg defects). Figure 4b shows DFTs of such maps; sharp non-dispersive Bragg peaks (at the corners, $(\pm 2\pi/a, 0), (0, \pm 2\pi/a)$) corresponding to the atomic lattice are seen, as well as other features (concentric square-like shapes) that rapidly disperse with energy, collapse (Fig. 4b; 0 meV) and then disappear (Fig. 4b; 9 meV) near the Fermi energy. We have carried out such measurements both at low temperatures (20 K, Fig. 4b), where the spectrum shows signatures of hybridization between conduction electrons and f orbitals, and at high temperatures (70 K, Fig. 4c), where such features are considerably weakened (for example, Fig. 4c; 2 meV, 10 meV). As a control experiment, we have also carried out the same measurements on the corresponding surface of $CeRhIn_5$ (Fig. 4d), for which signatures of heavy electron behaviour are absent (for example, Fig. 3a) in the same temperature window (20 K). Although understanding details of the QPI in Fig. 4 requires detailed modelling of the band structure of the 115 compounds, the square-like patterns observed in the data correspond to scattering wavevectors that can be identified from the calculated local-density approximation (LDA) band structure[46] (see Supplementary Information section V).

We find that analysing the features of the energy-resolved DFT maps provides direct evidence for mass enhancement of quasiparticles, in unison with related signatures in the tunnelling spectra. Figure 5a and b shows line sections of the DFT maps plotted along two high-symmetry directions (the thick white lines in Fig. 4b) as a function of energy for $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 20 K, and in Fig. 5c we show their

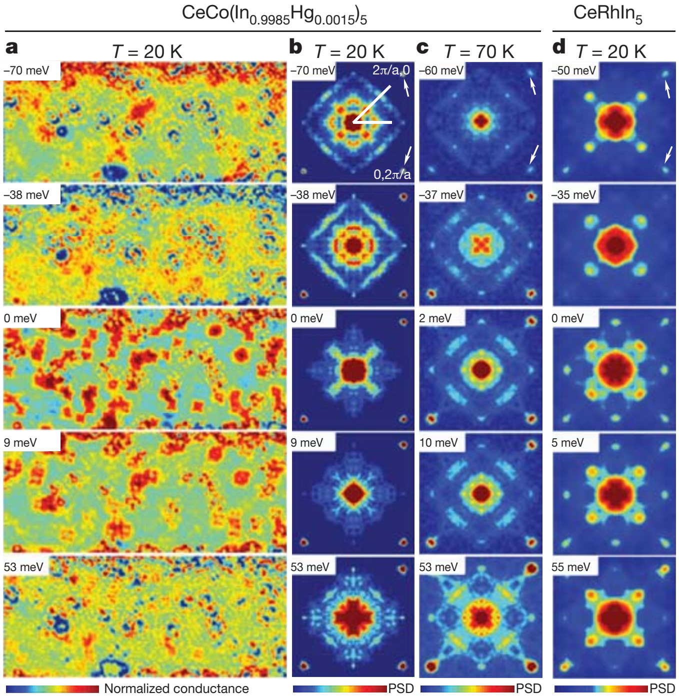


Figure 4 | Spectroscopic mapping of quasiparticle interference (QPI). **a**, **b**, Real space (**a**) and corresponding DFT (**b**) of conductance maps (−200 mV, 1.6 nA) at selected energies (top left of each panel) measured on surface A of $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 20 K. **c**, **d**, Similar DFTs for $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 70 K (−150 mV, 1.5 nA; **c**) and on the corresponding surface A for $CeRhIn_5$ at 20 K (−200 mV, 3.0 nA; **d**) at selected energies. Arrows indicate the position of the Bragg peaks at ($2\pi/a$, 0) and (0, $2\pi/a$). All DFTs were four-fold symmetrized (due to the four-fold crystal symmetry) to enhance resolution (see Supplementary Information section IV). The intensity is represented on a linear scale. PSD, power spectrum density.

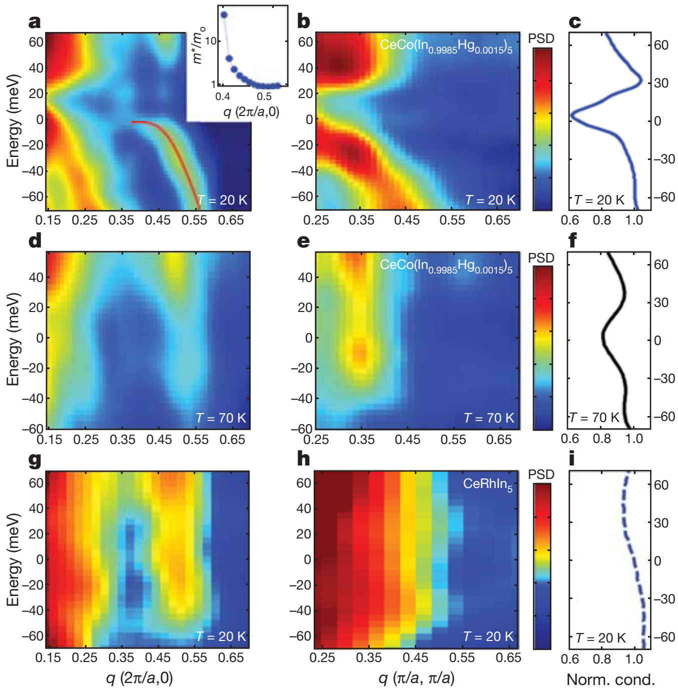


Figure 5 | Visualizing quasiparticle mass enhancement. **a**, **b**, Energy–momentum structure of the QPI bands extracted from line sections (solid white lines in Fig. 4b) along the atomic direction ($2\pi/a$, 0) (**a**) and along the zone diagonal (π/a, π/a) (**b**) in $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 20 K. The solid red line represents a fourth-order polynomial fit to the data. Inset in **a** shows the effective mass m^*/m_0 as a function of momenta obtained from the curvature ($\frac{1}{4}\hbar^2 (d^2E/dq^2)^{-1}$) of the outer band (solid red line in **a**). **c**, Average spectrum on surface A of $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 20 K, reflecting the suppression of scattering in the QPI bands. Similar measurements performed in $CeCo(In_{0.9985}Hg_{0.0015})_5$ at 70 K (**d**–**f**) and in $CeRhIn_5$ at 20 K (**g**–**i**). The intensity is represented on a linear scale.

corresponding spatially averaged spectrum. The square-like regions of enhanced quasiparticle scattering in Fig. 4b appear in the line sections of Fig. 5a, b as energy-dependent bands of scattering, which become strongly energy dependent near the Fermi energy. Clearly the scattering of the quasiparticle excitations in the energy window near the direct hybridization gap has a flatter energy–momentum structure than that at energies away from the gap. This is a direct signature of the quasiparticles acquiring heavy effective mass at low energies near the Fermi energy. Detailed analysis of one of the QPI bands estimates the mass enhancement near the Fermi energy to be about $30m_0$ (Fig. 5a inset), a value which is close to that seen in quantum oscillation studies of $CeCoIn_5$ (refs 39, 40). Our model calculation, which describes the spectroscopic lineshapes on the different surfaces, can also be extended to reproduce the signatures of mass enhancement in the QPI data (see Supplementary Information section I).

Contrasting low temperature QPI patterns on $CeCo(In_{0.9985}Hg_{0.0015})_5$ to measurements on the same compound at high temperatures (70 K, Fig. 5d, e), where the hybridization gap is weak (Fig. 5f), or to measurements on $CeRhIn_5$ (20 K, Fig. 5g, h), where signatures of a hybridization gap are absent in the tunnelling spectra (Fig. 5i), confirms that the development of this gap results in apparent splitting of the bands and is responsible for both the scattering and the heavy effective mass in the QPI measurements. Furthermore, these measurements show that the underlying band structure responsible for the scattering wavevectors away from the Fermi energy is relatively similar between $CeCo(In_{0.9985}Hg_{0.0015})_5$ and $CeRhIn_5$. Only when f electrons of the Kondo lattice begin to strongly hybridize with conduction electrons and modify the band structure within a relatively narrow energy window (30 meV) do we see signatures of heavy-fermion excitations in QPI measurements, signalling a transition from a small to a large Fermi surface (see Supplementary Information section V).

Signatures of quantum criticality

The ability to tunnel through the f component of the heavy quasiparticles on surface B of $CeCo(In_{0.9985}Hg_{0.0015})_5$ provides an opportunity to probe the lifetime of the heavy quasiparticles as a function of temperature in a system that is close to a QCP. The narrow dispersion of the f band results in a direct connection between the experimentally measured width of the peak in the density of states near the Fermi energy (Fig. 3b) and the lifetime of the heavy quasiparticles. Analysis of this width measured at different temperatures is displayed in Fig. 6a (see Supplementary Information section VI), and shows a strong temperature dependence with a finite intercept (~3.5 meV) in the limit of zero temperature. The finite width at zero temperature can be understood as a consequence of a small but finite dispersion of the f band as well as a finite probability of tunnelling into the *spd* electrons (see Supplementary Information section I). However, the large linear slope in Fig. 6a (that is, larger than $3/2\ k_BT$, where k_B is Boltzmann's constant and T is temperature) indicates that the lifetime of the f electrons, as opposed to thermal broadening, is strongly influencing the spectra and its temperature dependence. Consistent with this observation, we also find that to capture the temperature evolution of the spectra in Fig. 3b, we have to use rather large values of scattering rate (inverse lifetime) of the f component of the heavy quasiparticles, $\gamma_f = \hbar/\tau_f$ (where $\hbar$ is Planck's constant h divided by 2π and τ_f is the lifetime of the quasiparticles), in our model calculations (Fig. 3d).

A T-linear scattering rate (or inverse lifetime) for the heavy quasiparticles is consistent with the expectation that $CeCoIn_5$ is close to a QCP, because for systems tuned close to such transitions, temperature is the only relevant energy scale available to determine the quasiparticle lifetime, resulting in $\hbar/\tau_f \propto k_BT$ (refs 18, 47). However, a more precise signature of a QCP would be the observation

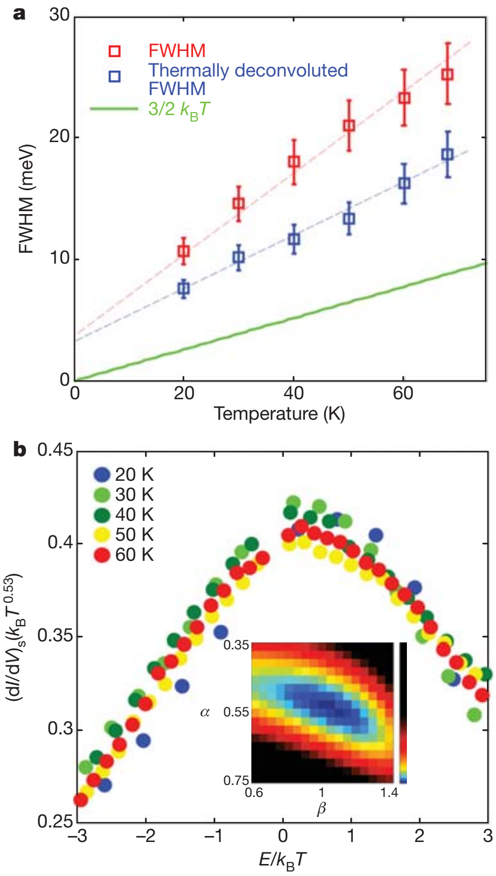


Figure 6 | Signatures of quantum criticality. **a**, Full width at half maximum (FWHM) of the heavy quasiparticle peak (red squares) as a function of temperature extracted from a Gaussian fit to the sharp lineshape of the spectra of Fig. 3b after a smooth background subtraction (see Supplementary Information section VI). Blue squares represent the thermally deconvoluted FWHM corresponding to the intrinsic width in the absence of thermal broadening. The green line represents $3/2k_BT$. Error bars, 1 s.d. **b**, Energy–temperature scaling of the different spectra of Fig. 3b, after the removal of a temperature independent background (see Supplementary Information sections VI and VII), within a narrow energy window near the Fermi energy. Key shows temperature T in K. Inset shows the 'goodness of the collapse' (colour scale) as a function of the critical exponents α and β.

of energy–temperature scaling of experimental quantities near such transitions. In fact, recent theoretical work suggests that the instability of the Fermi surface near a QCP should result in scaling properties of the single-particle excitation that can be directly probed in measurements of the tunnelling density of states[48]. To test this hypothesis, we examine the lineshape of the tunnelling spectra on surface B of $CeCo(In_{0.9985}Hg_{0.0015})_5$ near the chemical potential at different temperatures, and attempt to scale the data (Fig. 3b) by plotting $(dI/dV)_S(k_BT^{\alpha})$ as a function of $E/(k_BT)^{\beta}$. (Here $(dI/dV)_S$ is the background-subtracted spectra of Fig. 3b (see Supplementary Information section VI) and E is the energy of the tunnelling quasiparticles.) We find that using the exponents $\alpha = 0.53$ and $\beta = 1$ results in a collapse of the data at different temperatures on a single curve in the low bias region (see Fig. 6b and Supplementary Information section VII). Although an understanding of the value of the exponent α is currently lacking, the linear power β confirms our hypothesis of energy–temperature scaling associated with proximity to a QCP. These results indicate that the heavy quasiparticles in $CeCoIn_5$ are damped because of critical fluctuations rather than the typical scattering that is expected in a Fermi liquid (T^2 dependence). Similar energy–temperature scalings, with anomalous exponents (α), have been previously observed in the dynamical spin susceptibility of other heavy-fermion systems near QCPs[3,5,49,50]. However, here we show for the first time that the signatures of scaling and critical phenomena appear in the spectroscopic properties of the quasiparticle excitations.

Conclusion and outlook

The experimental results and the model calculations presented here provide a comprehensive picture of how heavy-fermion excitations in the 115 Ce-based Kondo lattice systems emerge with lowering of temperature or as a result of chemical tuning of the interaction between the Ce *f* electrons and the conduction electrons. The changes in the scattering properties of the quasiparticles directly signal the flattening of their energy–momentum structure and the emergence of heavy quasiparticles near the Fermi energy. Such changes are also consistent with the predicted evolution from a small to a large Fermi surface as the localized *f* electrons hybridize with the conduction electrons. The sensitivity of the tunnelling to the surface termination and the successful modelling of these data provide direct spectroscopic evidence of the composite nature of heavy fermions and offer a unique method to disentangle their components.

Our experiments also demonstrate that the emergent heavy quasiparticles in our system are strongly scattered and show signatures of scaling associated with critical damping of excitations in proximity to a QCP. Like many other heavy-fermion systems, thermodynamic and transport studies of the 115 systems have shown evidence of quantum criticality, but such signatures have not been previously isolated in an electron spectroscopy measurement, as described here. Such spectroscopic signatures are direct evidence for the breakdown of coherent fermionic excitations approaching a QCP. Future extension of our measurements to lower temperatures could probe the interplay between quantum fluctuations and the appearance of superconductivity, an issue which continues to be one of the most debated in condensed matter physics.

METHODS SUMMARY

The single crystals of $CeCoIn_5$, $CeCo(In_{0.9985}Hg_{0.0015})_5$ and $CeRhIn_5$ used for this study were grown from excess indium at Los Alamos National Laboratory. Small, flat crystals were oriented along the crystallographic axes and cut into sizes suitable for STM measurements ($\sim 2 \times 2 \times 0.2$ mm^3). The samples were cleaved on a surface perpendicular to the c axis in ultrahigh vacuum (UHV) and transferred *in situ* to the microscope head. Differential conductance (dI/dV) measurements were performed using standard lock-in techniques. Approximately ten different samples of $CeCoIn_5$, $CeCo(In_{0.9985}Hg_{0.0015})_5$ and $CeRhIn_5$ were successfully cleaved and studied, and the spectroscopic data collected were reproducible on the corresponding identical exposed surfaces of the different samples. Spectra measured at different locations on each surface showed negligible variations. The spectra presented here (in the main paper) are averaged over approximately 200 individual spectra measured over an area of at least 100 Å × 100 Å. The spectroscopic lineshapes showed negligible variations as the tip height was varied (variation of the tunnelling current by two orders of magnitude).

Received 21 December 2011; accepted 30 April 2012.

1. Anderson, P. W. Localized magnetic states in metals. *Phys. Rev.* **124,** 41–53 (1961).
2. Shiba, H& Kuramoto, Y. (eds) Kondo effect — 40 years after the discovery. *J. Phys. Soc. Jpn* **74,** 1–238 (2005).
3. Schroder, A. *et al.* Onset of antiferromagnetism in heavy-fermion metals. *Nature* **407,** 351–355 (2000).
4. Coleman, P., Pépin, C., Si, Q. & Ramazashvili, R. How do Fermi liquids get heavy and die? *J. Phys. Condens. Matter* **13,** 723–738 (2001).
5. Si, Q., Rabello, S., Ingersent, K. & Smith, J. L. Locally critical quantum phase transitions in strongly correlated metals. *Nature* **413,** 804–808 (2001).
6. Senthil, T., Sachdev, S. & Vojta, M. Fractionalized Fermi liquids. *Phys. Rev. Lett.* **90,** 216403 (2003).
7. Park, T. *et al.* Hidden magnetism and quantum criticality in the heavy fermion superconductor $CeRhIn_5$. *Nature* **440,** 65–68 (2006).
8. Gegenwart, P., Si, Q. & Steglich, F. Quantum criticality in heavy-fermion metals. *Nature Phys.* **4,** 186–197 (2008).
9. Pfleiderer, C. Superconducting phases of f-electron compounds. *Rev. Mod. Phys.* **81,** 1551–1624 (2009).
10. Palstra, T. T. M. *et al.* Superconducting and magnetic transitions in the heavy-fermion system URu_2Si_2. *Phys. Rev. Lett.* **55,** 2727–2730 (1985).
11. Stewart, G. Heavy-fermion systems. *Rev. Mod. Phys.* **56,** 755–787 (1984).

12. Fisk, Z., Sarrao, J. L., Smith, J. L. & Thompson, J. D. The physics and chemistry of heavy fermions. *Proc. Natl Acad. Sci. USA* **92,** 6663–6667 (1995).
13. Steglich, F. *et al.* Classification of strongly correlated f-electron systems. *J. Low Temp. Phys.* **99,** 267–281 (1995).
14. Yang, Y.-f., Fisk, Z., Lee, H.-O., Thompson, J. D. & Pines, D. Scaling the Kondo lattice. *Nature* **454,** 611–613 (2008).
15. Si, Q. & Steglich, F. Heavy fermions and quantum phase transitions. *Science* **329,** 1161–1166 (2010).
16. Anderson, P. W. Fermi sea of heavy electrons (a Kondo lattice) is never a Fermi liquid. *Phys. Rev. Lett.* **104,** 176403 (2010).
17. Coleman, P. in *Handbook of Magnetism and Advanced Magnetic Materials* Vol. 1 (eds Kronmeuller, H. & Parkin, S.) 45 (Wiley and Sons, 2007).
18. Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, 1999).
19. Varma, C. M. Mixed-valence compounds. *Rev. Mod. Phys.* **48,** 219–238 (1976).
20. Grenzebach, C., Anders, F. B., Czycholl, G. & Pruschke, T. Transport properties of heavy-fermion systems. *Phys. Rev. B* **74,** 195119 (2006).
21. Shim, J. H., Haule, K. & Kotliar, G. Modeling the localized-to-itinerant electronic transition in the heavy fermion system $CeIrIn_5$. *Science* **318,** 1615–1617 (2007).
22. Martin, L. C., Bercx, M. & Assaad, F. F. Fermi surface topology of the two-dimensional Kondo lattice model: dynamical cluster approximation approach. *Phys. Rev. B* **82,** 245105 (2010).
23. Jacob, D., Haule, K. & Kotliar, G. Dynamical mean-field theory for molecular electronics: electronic structure and transport properties. *Phys. Rev. B* **82,** 195115 (2010).
24. Benlagra, A., Pruschke, T. & Vojta, M. Finite-temperature spectra and quasiparticle interference in Kondo lattices: from light electrons to coherent heavy quasiparticles. *Phys. Rev. B* **84,** 195141 (2011).
25. Yang, Y.-f. Fano effect in the point contact spectroscopy of heavy-electron materials. *Phys. Rev. B* **79,** 241107 (2009).
26. Maltseva, M., Dzero, M. & Coleman, P. Electron cotunneling into a Kondo lattice. *Phys. Rev. Lett.* **103,** 206402 (2009).
27. Figgins, J. & Morr, D. K. Differential conductance and quantum interference in Kondo systems. *Phys. Rev. Lett.* **104,** 187202 (2010).
28. Wölfle, P., Dubi, Y. & Balatsky, A. V. Tunneling into clean heavy fermion compounds: origin of the Fano line shape. *Phys. Rev. Lett.* **105,** 246401 (2010).
29. Park, W. K., Sarrao, J. L., Thompson, J. D. & Greene, L. H. Andreev reflection in heavy-fermion superconductors and order parameter symmetry in $CeCoIn_5$. *Phys. Rev. Lett.* **100,** 177001 (2008).
30. Aynajian, P. *et al.* Visualizing the formation of the Kondo lattice and the hidden order in URu_2Si_2. *Proc. Natl Acad. Sci. USA* **107,** 10383–10388 (2010).
31. Schmidt, A. R. *et al.* Imaging the Fano lattice to 'hidden order' transition in URu_2Si_2. *Nature* **465,** 570–576 (2010).
32. Ernst, S. *et al.* Emerging local Kondo screening and spatial coherence in the heavy-fermion metal $YbRh_2Si_2$. *Nature* **474,** 362–366 (2011).
33. Sidorov, V. A. *et al.* Superconductivity and quantum criticality in $CeCoIn_5$. *Phys. Rev. Lett.* **89,** 157004 (2002).
34. Paglione, J. *et al.* Field-induced quantum critical point in $CeCoIn_5$. *Phys. Rev. Lett.* **91,** 246405 (2003).
35. Paglione, J., Sayles, T. A., Ho, P. C., Jeffries, J. R. & Maple, M. B. Incoherent non-Fermi-liquid scattering in a Kondo lattice. *Nature Phys.* **3,** 703–706 (2007).
36. Urbano, R. R. *et al.* Interacting antiferromagnetic droplets in quantum critical $CeCoIn_5$. *Phys. Rev. Lett.* **99,** 146402 (2007).
37. Hegger, H. *et al.* Pressure-induced superconductivity in quasi-2D $CeRhIn_5$. *Phys. Rev. Lett.* **84,** 4986–4989 (2000).
38. Petrovic, C. *et al.* Heavy-fermion superconductivity in $CeCoIn_5$ at 2.3 K. *J. Phys. Condens. Matter* **13,** 337–342 (2001).
39. Hall, D. *et al.* Fermi surface of the heavy-fermion superconductor $CeCoIn_5$: the de Haas–van Alphen effect in the normal state. *Phys. Rev. B* **64,** 212508 (2001).
40. Shishido, H., Settai, R., Hashimoto, S., Inada, Y. & Ōnuki, Y. De Hass van Alphen effect of $CeRhIn_5$ and $CeCoIn_5$ under pressure. *J. Magn. Magn. Mater.* **272–276,** 225–226 (2004).
41. Moshopoulou, E. G. *et al.* Comparison of the crystal structure of the heavy-fermion materials $CeCoIn_5$, $CeRhIn_5$ and $CeIrIn_5$. *Appl. Phys. A* **74,** s895–s897 (2002).
42. Booth, C. H. *et al.* Local structure and site occupancy of Cd and Hg substitutions in $CeTIn_5$ (T=Co, Rh, and Ir). *Phys. Rev. B* **79,** 144519 (2009).
43. Fujimori, S. *et al.* Direct observation of a quasiparticle band in $CeIrIn_5$: an angle-resolved photoemission spectroscopy study. *Phys. Rev. B* **73,** 224517 (2006).
44. Ehm, D. *et al.* High-resolution photoemission study on low-T_K Ce systems: Kondo resonance, crystal field structures, and their temperature dependence. *Phys. Rev. B* **76,** 045117 (2007).
45. Koitzsch, A. *et al.* Hybridization effects in $CeCoIn_5$ observed by angle-resolved photoemission. *Phys. Rev. B* **77,** 155128 (2008).
46. Oppeneer, P. M. *et al.* Fermi surface changes due to localized–delocalized f-state transitions in Ce-115 and Pu-115 compounds. *J. Magn. Magn. Mater.* **310,** 1684–1690 (2007).
47. Sachdev, S. & Ye, J. Universal quantum-critical dynamics of two-dimensional antiferromagnets. *Phys. Rev. Lett.* **69,** 2411–2414 (1992).
48. Senthil, T. Critical Fermi surfaces and non-Fermi liquid metals. *Phys. Rev. B* **78,** 035103 (2008).
49. Aronson, M. C. *et al.* Non-Fermi-liquid scaling of the magnetic response in $UCu_{5-x}Pd_x$ (x=1,1.5). *Phys. Rev. Lett.* **75,** 725–728 (1995).
50. Schröder, A., Aeppli, G., Bucher, E., Ramazashvili, R. & Coleman, P. Scaling of magnetic fluctuations near a quantum phase transition. *Phys. Rev. Lett.* **80,** 5623–5626 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge discussions with P. W. Anderson, E. Abrahams, P. Coleman, N. Curro, D. Pines, D. Morr, T. Senthil, S. Sachdev, M. Vojta, C. Varma and C. V. Parker. Work at Princeton University was primarily supported by a grant from the DOE Office of Basic Energy Sciences (DE-FG02-07ER46419). The instrumentation and infrastructure at the Princeton Nanoscale Microscopy Laboratory are also supported by grants from the NSF-DMR1104612 and NSF-MRSEC programmes through the Princeton Center for Complex Materials (DMR-0819860), and the W.M. Keck foundation as well as the Eric and Linda Schmidt Transformative fund at Princeton. P.A. acknowledges postdoctoral fellowship support through the Princeton Center for Complex Materials funded by the NSF-MRSEC programme. Work at Los Alamos National Laboratory was performed under the auspices of the US Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering. Z.F. acknowledges support from NSF-DMR-0801253.

Author Contributions P.A., E.H.d.S.N. and A.G. performed the STM measurements. P.A. and E.H.d.S.N. analysed the data. E.H.d.S.N. and P.A. performed the theoretical calculations. R.E.B., J.D.T., Z.F. and E.D.B. synthesized and characterized the materials. A.Y., P.A. and E.H.d.S.N. wrote the manuscript. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.Y. (yazdani@princeton.edu).

doi:10.1038/nature11234

Structure, function and diversity of the healthy human microbiome

The Human Microbiome Project Consortium*

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81–99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

A total of 4,788 specimens from 242 screened and phenotyped adults[1] (129 males, 113 females) were available for this study, representing the majority of the target Human Microbiome Project (HMP) cohort of 300 individuals. Adult subjects lacking evidence of disease were recruited based on a lengthy list of exclusion criteria; we will refer to them here as 'healthy', as defined by the consortium clinical sampling criteria (K. Aagaard *et al.*, manuscript submitted). Women were sampled at 18 body habitats, men at 15 (excluding three vaginal sites), distributed among five major body areas. Nine specimens were collected from the oral cavity and oropharynx: saliva; buccal mucosa (cheek), keratinized gingiva (gums), palate, tonsils, throat and tongue soft tissues, and supra- and subgingival dental plaque (tooth biofilm above and below the gum). Four skin specimens were collected from the two retroauricular creases (behind each ear) and the two antecubital fossae (inner elbows), and one specimen for the anterior nares (nostrils). A self-collected stool specimen represented the microbiota of the lower gastrointestinal tract, and three vaginal specimens were collected from the vaginal introitus, midpoint and posterior fornix. To evaluate within-subject stability of the microbiome, 131 individuals in these data were sampled at an additional time point (mean 219 days and s.d. 69 days after first sampling, range 35–404 days). After quality control, these specimens were used for 16S rRNA gene analysis via 454 pyrosequencing (abbreviated henceforth as 16S profiling, mean 5,408 and s.d. 4,605 filtered sequences per sample); to assess function, 681 samples were sequenced using paired-end Illumina shotgun metagenomic reads (mean 2.9 gigabases (Gb) and s.d. 2.1 Gb per sample)[1]. More details on data generation are provided in related HMP publications[1] and in Supplementary Methods.

Microbial diversity of healthy humans

The diversity of microbes within a given body habitat can be defined as the number and abundance distribution of distinct types of organisms, which has been linked to several human diseases: low diversity in the gut to obesity and inflammatory bowel disease[2,3], for example, and high diversity in the vagina to bacterial vaginosis[4]. For this large study involving microbiome samples collected from healthy volunteers at two distinct geographic locations in the United States, we have defined the microbial communities at each body habitat, encountering 81–99% of predicted genera and saturating the range of overall community configurations (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1; see also Fig. 4). Oral and stool communities were especially diverse in terms of community membership, expanding prior observations[5], and vaginal sites harboured particularly simple communities (Fig. 1a). This study established that these patterns of alpha diversity (within samples) differed markedly from comparisons between samples from the same habitat among subjects (beta diversity, Fig. 1b). For example, the saliva had among the highest median alpha diversities of operational taxonomic units (OTUs, roughly species level classification, see http://hmpdacc.org/HMQCP), but one of the lowest beta diversities—so although each individual's saliva was ecologically rich, members of the population shared similar organisms. Conversely, the antecubital fossae (skin) had the highest beta diversity but were intermediate in alpha diversity. The vagina had the lowest alpha diversity, with quite low beta diversity at the genus level but very high among OTUs due to the presence of distinct *Lactobacillus* spp. (Fig. 1b). The primary patterns of variation in community structure followed the major body habitat groups (oral, skin, gut and vaginal), defining as a result the complete range of population-wide between-subject variation in human microbiome habitats (Fig. 1c). Within-subject variation over time was consistently lower than between-subject variation, both in organismal composition and in metabolic function (Fig. 1d). The uniqueness of each individual's microbial community thus seems to be stable over time (relative to the population as a whole), which may be another feature of the human microbiome specifically associated with health.

No taxa were observed to be universally present among all body habitats and individuals at the sequencing depth employed here, unlike several pathways (Fig. 2 and Supplementary Fig. 2, see below), although several clades demonstrated broad prevalence and relatively abundant carriage patterns[6,7]. Instead, as suggested by individually

*Lists of participants and their affiliations appear at the end of the paper.

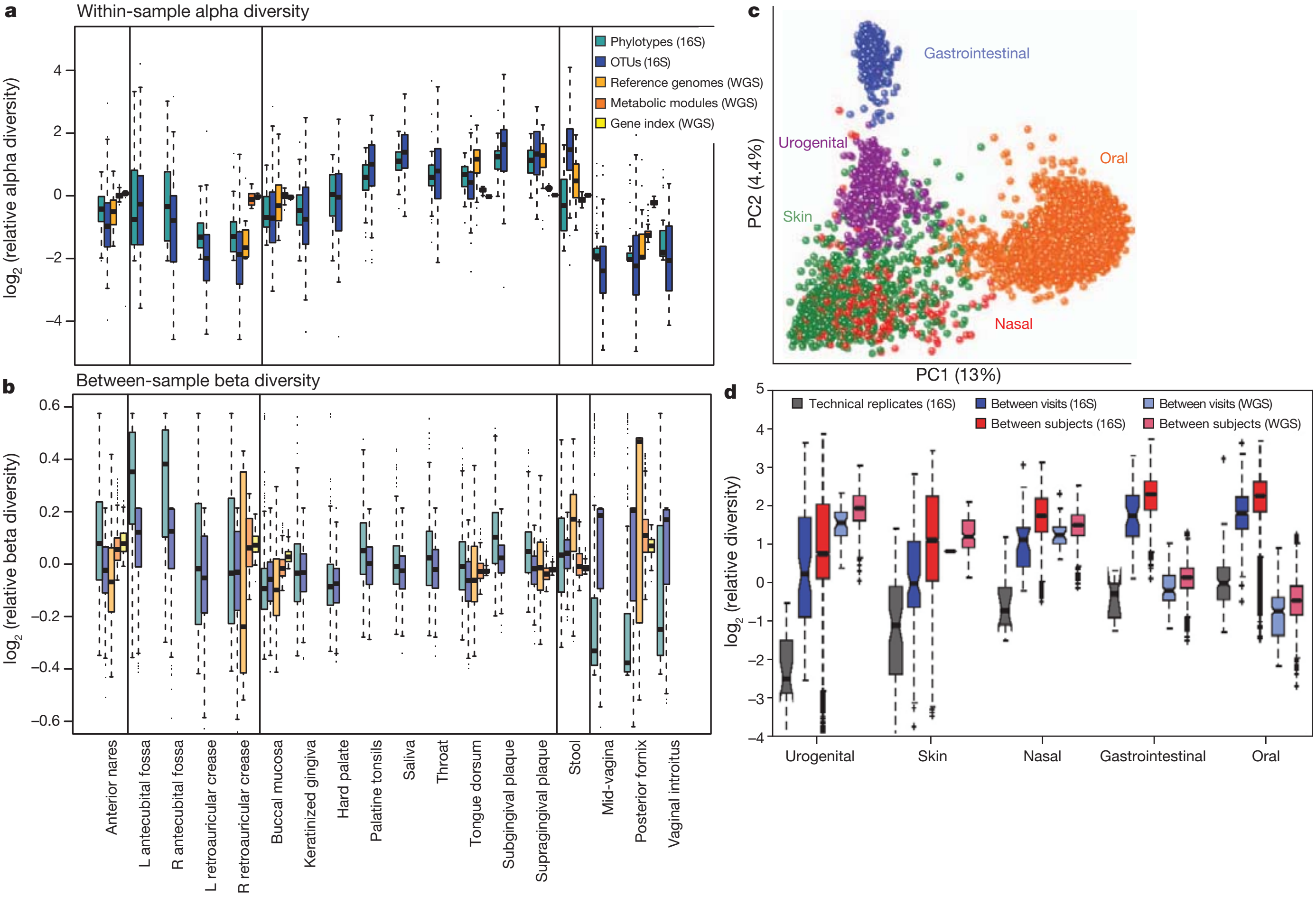


Figure 1 | Diversity of the human microbiome is concordant among measures, unique to each individual, and strongly determined by microbial habitat. **a**, Alpha diversity within subjects by body habitat, grouped by area, as measured using the relative inverse Simpson index of genus-level phylotypes (cyan), 16S rRNA gene OTUs (blue), shotgun metagenomic reads matched to reference genomes (orange), functional modules (dark orange), and enzyme families (yellow). The mouth generally shows high within-subject diversity and the vagina low diversity, with other habitats intermediate; variation among individuals often exceeds variation among body habitats. **b**, Bray–Curtis beta diversity among subjects by body habitat, colours as for **a**. Skin differs most between subjects, with oral habitats and vaginal genera more stable. Although alpha- and beta-diversity are not directly comparable, changes in structure among communities (**a**) occupy a wider dynamic range than do changes within communities among individuals (**b**). **c**, Principal coordinates plot showing variation among samples demonstrates that primary clustering is by body area, with the oral, gastrointestinal, skin and urogenital habitats separate; the nares habitat bridges oral and skin habitats. **d**, Repeated samples from the same subject (blue) are more similar than microbiomes from different subjects (red). Technical replicates (grey) are in turn more similar; these patterns are consistent for all body habitats and for both phylogenetic and metabolic community composition. See previously described sample counts[1] for all comparisons.

focused studies[2,3,5,8,9], each body habitat in almost every subject was characterized by one or a few signature taxa making up the plurality of the community (Fig. 3). Signature clades at the genus level formed on average anywhere from 17% to 84% of their respective body habitats, completely absent in some communities (0% at this level of detection) and representing the entire population (100%) in others. Notably, less dominant taxa were also highly personalized, both among individuals and body habitats; in the oral cavity, for example, most habitats are dominated by *Streptococcus*, but these are followed in abundance by *Haemophilus* in the buccal mucosa, *Actinomyces* in the supragingival plaque, and *Prevotella* in the immediately adjacent (but low oxygen) subgingival plaque[10].

Additional taxonomic detail of the human microbiome was provided by identifying unique marker sequences in metagenomic data[11] (Fig. 3a) to complement 16S profiling (Fig. 3b). These two profiles were typically in close agreement (Supplementary Fig. 3), with the former in some cases offering more specific information on members of signature genera differentially present within habitats (for example, vaginal *Prevotella amnii* and gut *Prevotella copri*) or among individuals (for example, vaginal *Lactobacillus* spp.) One application of this specificity was to confirm the absence of NIAID (National Institute of Allergy and Infectious Diseases) class A–C pathogens above 0.1% abundance (aside from *Staphylococcus aureus* and *Escherichia coli*) from the healthy microbiome, but the near-ubiquity and broad distribution of opportunistic 'pathogens' as defined by PATRIC[12]. Canonical pathogens including *Vibrio cholerae*, *Mycobacterium avium*, *Campylobacter jejuni* and *Salmonella enterica* were not detected at this level of sensitivity. *Helicobacter pylori* was found in only two stool samples, both at <0.01%, and *E. coli* was present at >0.1% abundance in 15% of stool microbiomes (>0% abundance in 61%). Similar species-level observations were obtained for a small subset of stool samples with 454 pyrosequencing metagenomics data using PhylOTU[13,14]. In total 56 of 327 PATRIC pathogens were detected in the healthy microbiome (at >1% prevalence of >0.1% abundance, Supplementary Table 2), all opportunistic and, strikingly, typically prevalent both among hosts and habitats. The latter is in contrast to many of the most abundant signature taxa, which were usually more habitat-specific and variable among hosts (Fig. 3a, b). This overall absence of particularly detrimental microbes supports the hypothesis that even given this cohort's high diversity, the microbiota tend to occupy a range of configurations in health distinct from many of the disease perturbations studied to date[3,15].

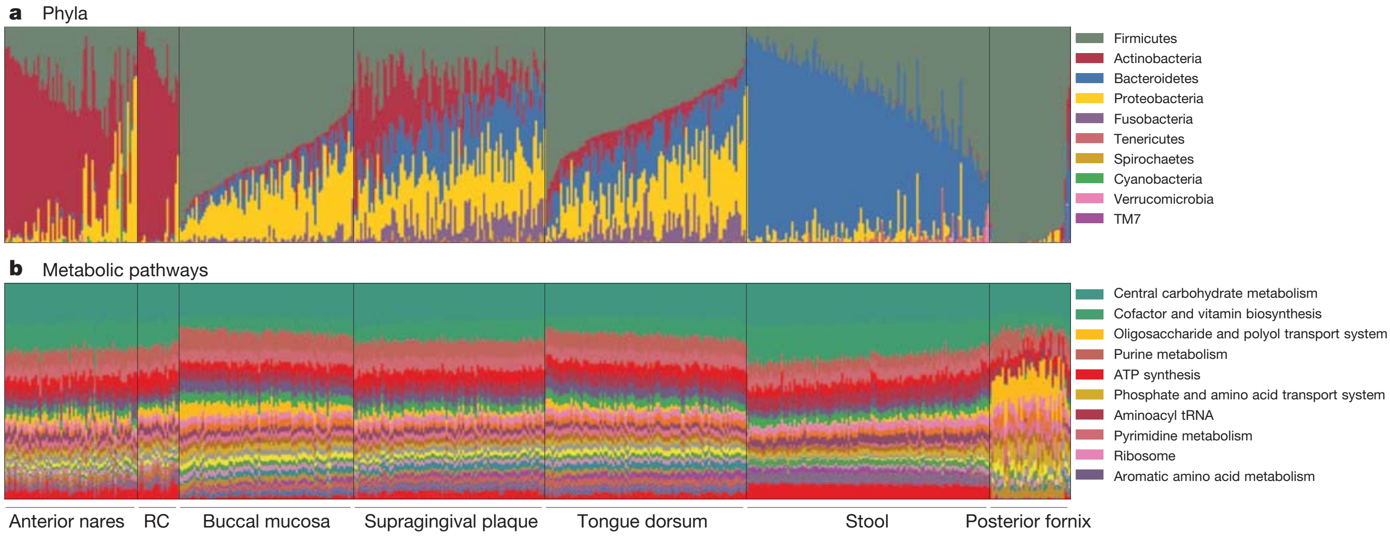


Figure 2 | Carriage of microbial taxa varies while metabolic pathways remain stable within a healthy population. **a**, **b**, Vertical bars represent microbiome samples by body habitat in the seven locations with both shotgun and 16S data; bars indicate relative abundances colored by microbial phyla from binned OTUs (**a**) and metabolic modules (**b**). Legend indicates most abundant phyla/pathways by average within one or more body habitats; RC, retroauricular crease. A plurality of most communities' memberships consists of a single dominant phylum (and often genus; see Supplementary Fig. 2), but this is universal neither to all body habitats nor to all individuals. Conversely, most metabolic pathways are evenly distributed and prevalent across both individuals and body habitats.

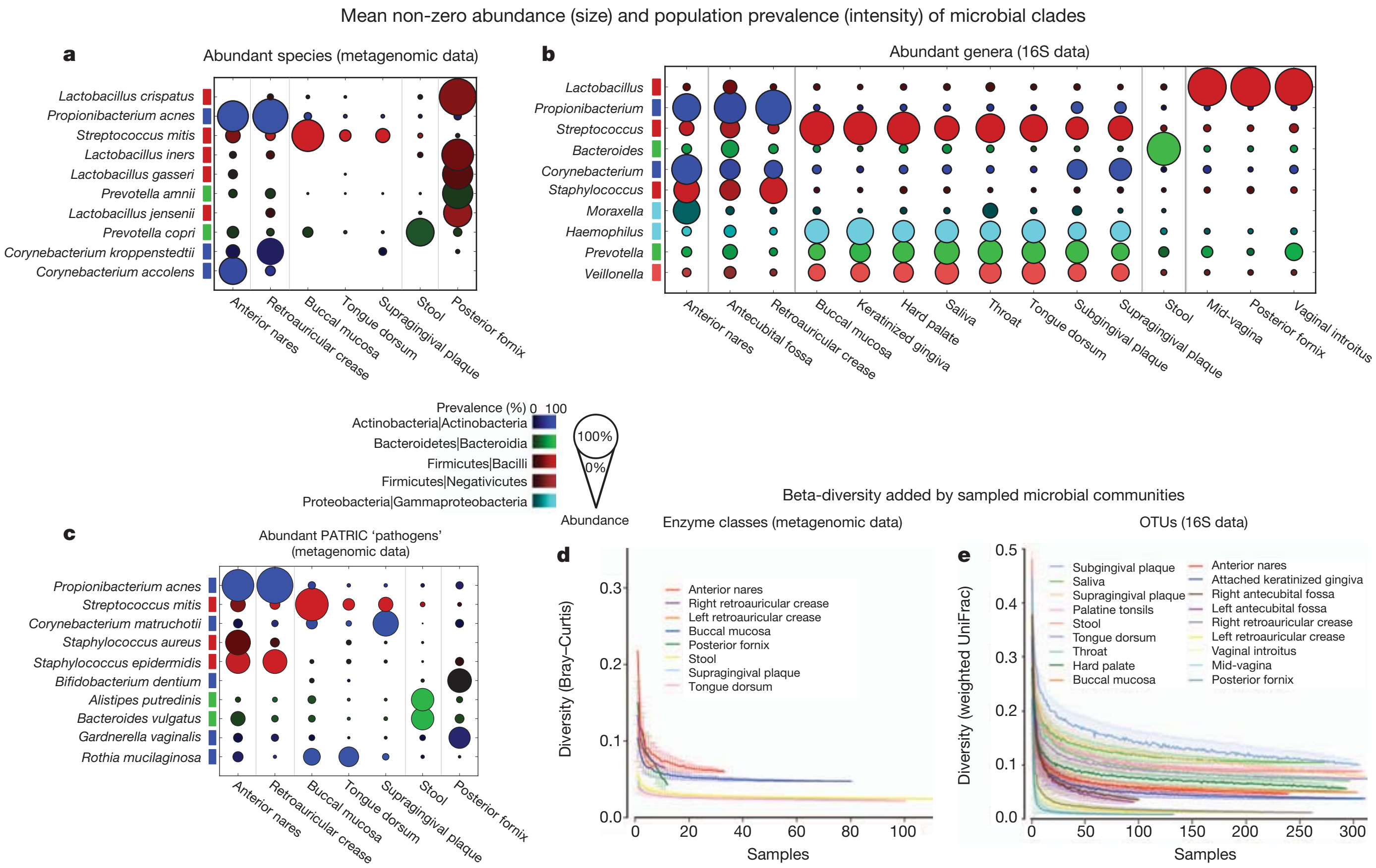


Figure 3 | Abundant taxa in the human microbiome that have been metagenomically and taxonomically well defined in the HMP population. **a**–**c**, Prevalence (intensity, colour denoting phylum/class) and abundance when present (size) of clades in the healthy microbiome. The most abundant metagenomically-identified species (**a**), 16S-identified genera (**b**) and PATRIC[12] pathogens (metagenomic) (**c**) are shown. **d**, **e**, The population size and sequencing depths of the HMP have well defined the microbiome at all assayed body sites, as assessed by saturation of added community metabolic configurations (rarefaction of minimum Bray–Curtis beta-diversity of metagenomic enzyme class abundances to nearest neighbour, inter-quartile range over 100 samples) (**d**) and phylogenetic configurations (minimum 16S OTU weighted UniFrac distance to nearest neighbour) (**e**).

Carriage of specific microbes

Inter-individual variation in the microbiome proved to be specific, functionally relevant and personalized. One example of this is illustrated by the *Streptococcus* spp. of the oral cavity. The genus dominates the oropharynx[16], with different species abundant within each sampled body habitat (see http://hmpdacc.org/HMSMCP) and, even at the species level, marked differences in carriage within each habitat among individuals (Fig. 4a). As the ratio of pan- to core-genomes is high in many human-associated microbes[17], this variation in abundance could be due to selective pressures acting on pathways differentially present among *Streptococcus* species or strains (Fig. 4b). Indeed, we observed extensive strain-level genomic variation within microbial species in this population, enriched for host-specific structural variants around genomic islands (Fig. 4c). Even with respect to the single *Streptococcus mitis* strain B6, gene losses associated with these events were common, for example differentially eliminating *S. mitis* carriage of the V-type ATPase or choline binding proteins cbp6 and cbp12 among subsets of the host population (Fig. 4d). These losses were easily observable by comparison to reference isolate genomes, and these initial findings indicate that microbial strain- and host-specific gene gains and polymorphisms may be similarly ubiquitous.

Other examples of functionally relevant inter-individual variation at the species and strain levels occurred throughout the microbiome. In the gut, *Bacteroides fragilis* has been shown to prime T-cell responses in animal models via the capsular polysaccharide A[18], and in the HMP stool samples this taxon was carried at a level of at least 0.1% in 16% of samples (over 1% abundance in 3%). *Bacteroides thetaiotaomicron* has been studied for its effect on host gastrointestinal metabolism[19] and was likewise common at 46% prevalence. On the skin, *S. aureus*, of particular interest as the cause of methicillin-resistant

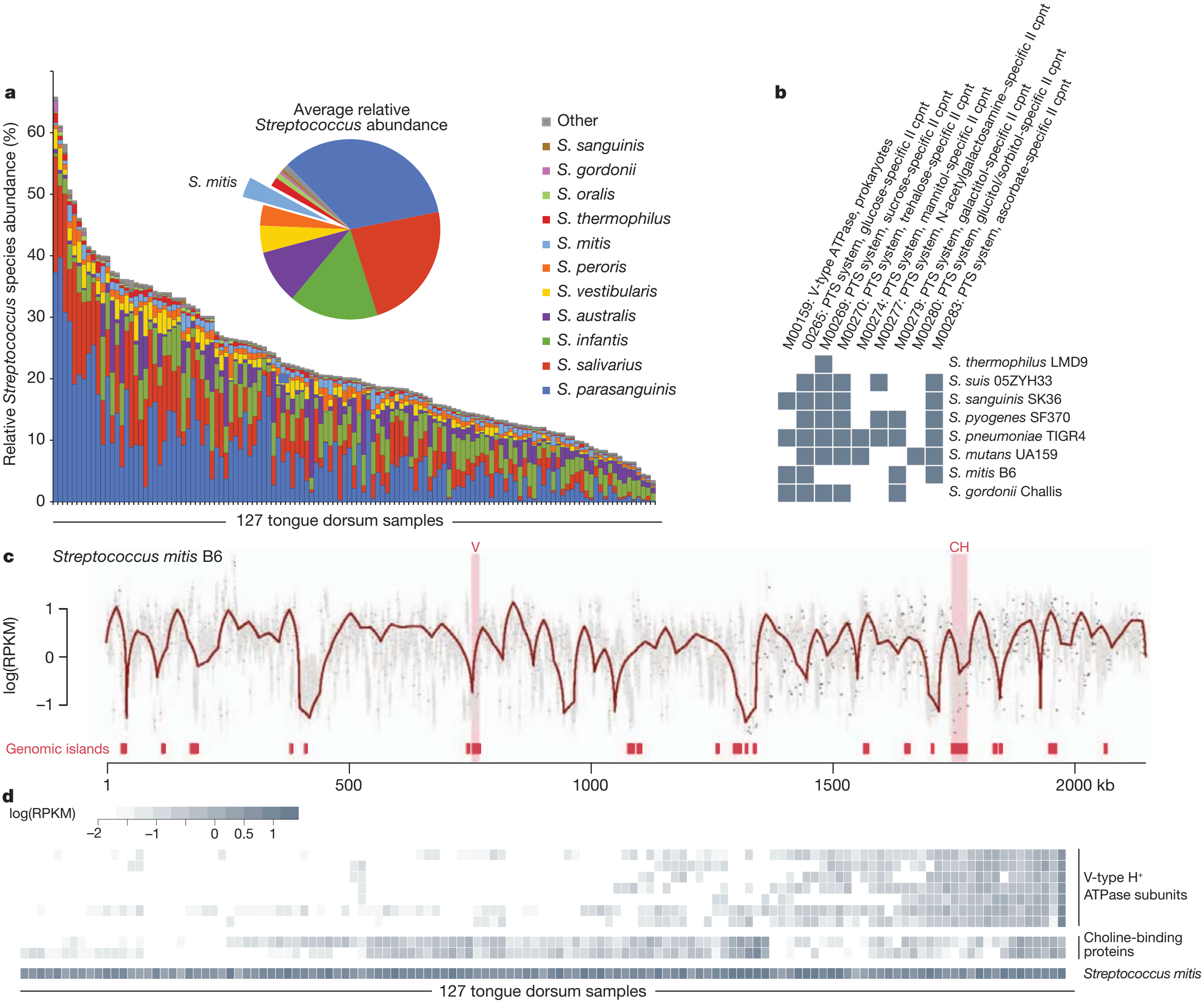


Figure 4 | Microbial carriage varies between subjects down to the species and strain level. Metagenomic reads from 127 tongue samples spanning 90 subjects were processed with MetaPhlAn to determine relative abundances for each species. **a**, Relative abundances of 11 distinct *Streptococcus* spp. In addition to variation in broader clades (see Fig. 2), individual species within a single habitat demonstrate a wide range of compositional variation. Inset illustrates average tongue sample composition. **b**, Metabolic modules present/absent (grey/white) in KEGG[24] reference genomes of tongue streptococci denote selected areas of strain-specific functional differentiation. cpnt, component. **c**, Comparative genomic coverage for the single *Streptococcus mitis* B6 strain. Grey dots are median reads per kilobase per million reads (RPKM) for 1-kb windows, grey bars are the 25th to 75th percentiles across all samples, red line the LOWESS-smoothed average. Red bars at the bottom highlight predicted genomic islands[27]. Large, discrete, and highly variable islands are commonly under-represented. **d**, Two islands are highlighted, V (V-type H^+ ATPase subunits I, K, E, C, F, A and B) and CH (choline-binding proteins cbp6 and cbp12), indicating functional cohesion of strain-specific gene loss within individual human hosts.

S. aureus (MRSA) infections, had 29% nasal and 4% skin carriage rates, roughly as expected[20]. Close phylogenetic relatives such as *Staphylococcus epidermidis* (itself considered commensal) were, in contrast, universal on the skin and present in 93% of nares samples, and at the opposite extreme *Pseudomonas aeruginosa* (a representative Gram-negative skin pathogen) was completely absent from both body habitats (0% at this level of detection). These and the data above suggest that the carriage pattern of some species in the human microbiome may be analogous to genetic traits, where recessive alleles of modest risk are maintained in a population. In the case of the human microbiome, high-risk pathogens remain absent, whereas species that pose a modest degree of risk also seem to be stably maintained in this ecological niche.

Finally, microorganisms within and among body habitats exhibited relationships suggestive of driving physical factors such as oxygen, moisture and pH, host immunological factors, and microbial interactions such as mutualism or competition[21] (Supplementary Fig. 4). Both overall community similarity and microbial co-occurrence and co-exclusion across the human microbiome grouped the 18 body habitats together into four clusters corresponding to the five target body areas (Supplementary Fig. 4a, b). There was little distinction among different vaginal sites, with *Lactobacillus* spp. dominating all three and correlating in abundance. However, *Lactobacillus* varied inversely with the Actinobacteria and Bacteroidetes (see Supplementary Fig. 4c and Figs 2 and 3), as also observed in a previous cohort[9]. Gut microbiota relationships primarily comprised inverse associations with the *Bacteroides*, which ranged from dominant in some subjects to a minority in others who carried a greater diversity of Firmicutes. A similar progression was evident in the skin communities, dominated by one of *Staphylococcus* (phylum Firmicutes), *Propionibacterium*, or *Corynebacterium* (both phylum Actinobacteria), with a continuum of oral organisms (for example, *Streptococcus*) appearing in nares communities (Supplementary Fig. 4c). These observations suggest that microbial community structure in these individuals may sometimes occupy discrete configurations and under other circumstances vary continuously, a topic addressed in more detail by several HMP investigations (ref. 6 and unpublished results). An individual's location within such configurations is indicative of current microbial carriage (including pathogens) and of the community's ability to resist future pathogen acquisition or dysbiosis; it may thus prove to be associated with disease susceptibility or other phenotypic characteristics.

Microbiome metabolism and function

As the first study to include both marker gene and metagenomic data across body habitats from a large human population, we additionally assessed the ecology of microbial metabolic and functional pathways in these communities. We reconstructed the relative abundances of pathways in community metagenomes[22], which were much more constant and evenly diverse than were organismal abundances (Fig. 2b, see also Fig. 1), confirming this as an ecological property of the entire human microbiome[2]. We were likewise able to determine for the first time that taxonomic and functional alpha diversity across microbial communities significantly correlate (Spearman of inverse Simpson's $r = 0.60$, $P = 3.6 \times 10^{-67}$, $n = 661$), the latter within a more proscribed range of community configurations (Supplementary Fig. 5).

Unlike microbial taxa, several pathways were ubiquitous among individuals and body habitats. The most abundant of these 'core' pathways include the ribosome and translational machinery, nucleotide charging and ATP synthesis, and glycolysis, and reflect the basics of host-associated microbial life. Also in contrast to taxa, few pathways were highly variable among subjects within any body habitat; exceptions included the Sec (orally, pathway relative abundance s.d. = 0.0052; total mean of oral standard deviations = 0.0011 with s.d. = 0.0016) and Tat (globally, pathway s.d. = 0.0055; mean of global standard deviations = 0.0023 with s.d. = 0.0033) secretion systems, indicating a high degree of host–microbe and microbe–microbe interactions in the healthy human microbiota. This high variability was particularly present in the oral cavity; for phosphate, mono- and di-saccharide, and amino acid transport in the mucosa; and also for lipopolysaccharide biosynthesis and spermidine/putrescine synthesis and transport on the plaque and tongue (http://hmpdacc.org/HMMRC). The stability and high metagenomic abundance of this housekeeping 'core' contrasts with the greater variability and lower abundance of niche-specific functionality in rare but consistently present pathways; for example, spermidine biosynthesis, methionine degradation and hydrogen sulphide production, all examples highly prevalent in gastrointestinal body sites (non-zero in >92% of samples) but at very low abundance (median relative abundance < 0.0052). This 'long tail' of low-abundance genes and pathways also probably encodes much of the uncharacterized biomolecular function and metabolism of these metagenomes, the expression levels of which remain to be explored in future metatranscriptomic studies.

Protein families showed diversity and prevalence trends similar to those of full pathways, ranging from maxima of only ~16,000 unique families per community in the vagina to almost 400,000 in the oral cavity (Fig. 1a, b; http://hmpdacc.org/HMGI). A remarkable fraction of these families were indeed functionally uncharacterized, including those detected by read mapping, with a minimum in the oral cavity (mean 58% s.d. 6.8%) and maximum in the nares (mean 77% s.d. 11%). Likewise, many genes annotated from assemblies could not be assigned a metabolic function, with a minimum in the vagina (mean 78% s.d. 3.4%) and maximum in the gut (mean 86% s.d. 0.9%). The latter range did not differ substantially by body habitat and is in close agreement with previous comprehensive gene catalogues of the gut metagenome[3]. Taken together with the microbial variation observed above throughout the human microbiome, functional variation among individuals might indicate pathways of particular importance in maintaining community structure in the face of personalized immune, environmental or dietary exposures among these subjects. Determining the functions of uncharacterized core and variable protein families will be especially essential in understanding role of the microbiota in health and disease.

Correlations with host phenotype

We finally examined relationships associating both clades and metabolism in the microbiota with host properties such as age, gender, body mass index (BMI), and other available clinical metadata (Fig. 5 and Supplementary Table 3). Using a sparse multivariate model, 960 microbial, enzymatic or pathway abundances were significantly associated with one or more of 15 subject phenotype and sample metadata features. A wide variety of taxa, gene families and metabolic pathways were differentially distributed with subject ethnicity at every body habitat (Fig. 5a), representing the phenotype with the greatest number (266 at false discovery rate (FDR) $q < 0.2$) of total associations with the microbiome. Vaginal pH has also been observed to correlate with microbiome composition[9], and we detected in this population both the expected reduction in *Lactobacillus* at high pH and a corresponding increase in metabolic diversity (Fig. 5b). Intriguingly, and not previously observed, subject age was most associated with a collection of highly differential metagenomically encoded pathways on the skin (Fig. 5c), as well as shifts in skin clades including retroauricular Firmicutes ($P = 1.0 \times 10^{-4}$, $q = 0.033$). The examples of associations with ethnicity and vaginal pH are among the strongest associations with the microbiome, however, and most correlates (for example, with subject BMI, Fig. 5d) are more representatively modest. This lower degree of correlation held for most available biometrics (gender, temperature, blood pressure, etc.), with even the most significant associations possessing generally low effect sizes and considerable unexplained variance. We conclude that most variation in the human microbiome is not well explained by these phenotypic

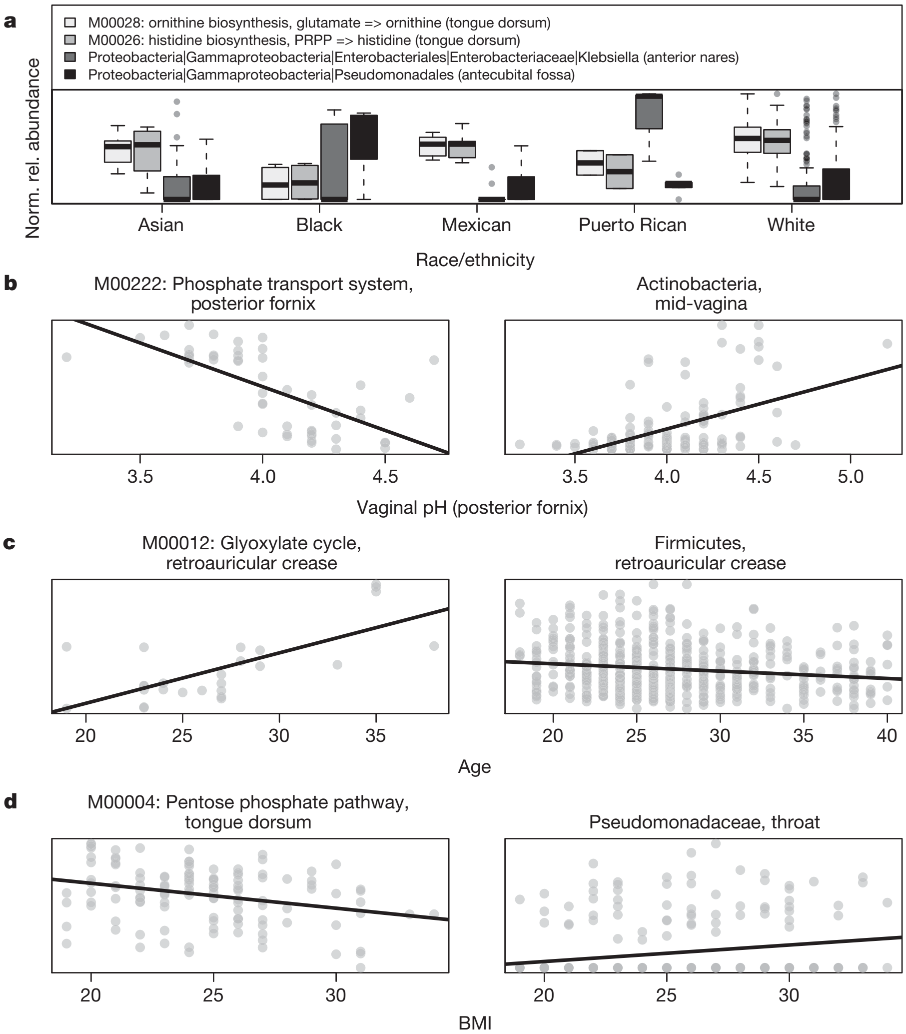


Figure 5 | Microbial community membership and function correlates with host phenotype and sample metadata. a–d, The pathway and clade abundances most significantly associated (all FDR $q < 0.2$) using a multivariate linear model with subject race or ethnicity (**a**), vaginal posterior fornix pH (**b**), subject age (**c**) and BMI (**d**). Scatter plots of samples are shown with lines indicating best simple linear fit. Race/ethnicity and vaginal pH are particularly strong associations; age and BMI are more representative of typically modest phenotypic associations (Supplementary Table 3), suggesting that variation in the healthy microbiota may correspond to other host or environmental factors.

metadata, and other potentially important factors such as short- and long-term diet, daily cycles, founder effects such as mode of delivery, and host genetics should be considered in future analyses.

Conclusions

This extensive sampling of the human microbiome across many subjects and body habitats provides an initial characterization of the normal microbiota of healthy adults in a Western population. The large sample size and consistent sampling of many sites from the same individuals allows for the first time an understanding of the relationships among microbes, and between the microbiome and clinical parameters, that underpin the basis for individual variation—variation that may ultimately be critical for understanding microbiome-based disorders. Clinical studies of the microbiome will be able to leverage the resulting extensive catalogues of taxa, pathways and genes[1], although they must also still include carefully matched internal controls. The uniqueness of each individual's microbiome even in this reference population argues for future studies to consider prospective within-subjects designs where possible. The HMP's unique combination of organismal and functional data across body habitats, encompassing both 16S and metagenomic profiling, together with detailed characterization of each subject, has allowed us and subsequent studies to move beyond the observation of variability in the human microbiome to ask how and why these microbial communities vary so extensively.

Many details remain for further work to fill in, building on this reference study. How do early colonization and lifelong change vary among body habitats? Do epidemiological patterns of transmission of beneficial or harmless microbes mirror patterns of transmission of pathogens? Which co-occurrences among microbes reflect shared response to the environment, as opposed to competitive or mutualistic interactions? How large a role does host immunity or genetics play in shaping patterns of diversity, and how do the patterns observed in this North American population compare to those around the world? Future studies building on the gene and organism catalogues established by the Human Microbiome Project, including increasingly detailed investigations of metatranscriptomes and metaproteomes, will help to unravel these open questions and allow us to more fully understand the links between the human microbiome, health and disease.

METHODS SUMMARY

Microbiome samples were collected from up to 18 body sites at one or two time points from 242 individuals clinically screened for absence of disease (K. Aagaard *et al.*, manuscript submitted). Samples were subjected to 16S ribosomal RNA gene pyrosequencing (454 Life Sciences), and a subset were shotgun-sequenced for metagenomics using the Illumina GAIIx platform[1]. 16S data processing and

diversity estimates were performed using QIIME[23], and metagenomic data were taxonomically profiled using MetaPhlAn[11], metabolically profiled by HUMAnN[22], and assembled for gene annotation and clustering into a unique catalogue[1]. Potential pathogens were identified using the PATRIC database[12], isolate reference genome annotations drawn from KEGG[24], and reference genome mapping performed by BWA[25] to a reduced set of genomes to which short reads could be matched[26]. Microbial associations were assessed by similarity measures accounting for compositionality[21], and phenotypic association testing was performed in R. All data and additional protocol details are available at http://hmpdacc.org. Full methods accompany this paper in the Supplementary Information.

Received 2 November 2011; accepted 16 May 2012.

1. The Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* http://dx.doi.org/10.1038/nature11209 (this issue).
2. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
3. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
4. Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular identification of bacteria associated with bacterial vaginosis. *N. Engl. J. Med.* **353,** 1899–1911 (2005).
5. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–1697 (2009).
6. Huse, S., Ye, Y., Zhou, Y. & Fodor, A. A core human microbiome as viewed through 16s rRNA sequences clusters. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.0034242 (14 June 2012).
7. Li, K., Bihan, M., Yooseph, S. & Methe, B. A. Analyses of the microbial diversity across the human microbiome. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.0032118 (14 June 2012).
8. Grice, E. A. *et al.* Topographical and temporal diversity of the human skin microbiome. *Science* **324,** 1190–1192 (2009).
9. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4680–4687 (2011).
10. Segata, N. *et al.* Composition of the adult digestive tract microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol.* **13,** R42 (2012).
11. Segata, N. *et al.* Efficient metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* http://dx.doi.org/10.1038/nmeth.2066 (2012).
12. Gillespie, J. J. *et al.* PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.* **79,** 4286–4298 (2011).
13. Sharpton, T. J. *et al.* PhylOTU: a high-throughput procedure quantifies microbial community diversity and resolves novel taxa from metagenomic data. *PLoS Comput. Biol.* **7,** e1001061 (2011).
14. Wylie, K. M. *et al.* Novel bacterial taxa in the human microbiome. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.003529 (14 June 2012).
15. Sokol, H. *et al. Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105,** 16731–16736 (2008).
16. Aas, J. A., Paster, B. J., Stokes, L. N., Olsen, I. & Dewhirst, F. E. Defining the normal bacterial flora of the oral cavity. *J. Clin. Microbiol.* **43,** 5721–5732 (2005).
17. Medini, D. *et al.* Microbiology in the post-genomic era. *Nature Rev. Microbiol.* **6,** 419–430 (2008).
18. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453,** 620–625 (2008).
19. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6,** 279–289 (2009).
20. Kuehnert, M. J. *et al.* Prevalence of *Staphylococcus aureus* nasal colonization in the United States, 2001–2002. *J. Infect. Dis.* **193,** 172–179 (2006).
21. Faust, K. *et al.* Microbial co-occurrence relationships in the human microbiome. *PLoS Comput. Biol.*. (in the press).
22. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* http://dx.doi.org/10.1371/journal.pcbi.1002358 (14 June 2012).
23. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
24. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38,** D355–D360 (2010).
25. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26,** 589–595 (2010).
26. Giannoukos, G. *et al.* Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* **13,** R23 (2012).
27. Langille, M. G. & Brinkman, F. S. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25,** 664–665 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Consortium would like to thank our external scientific advisory board: R. Blumberg, J. Davies, R. Holt, P. Ossorio, F. Ouellette, G. Schoolnik and A. Williamson. We would also like to thank our collaborators throughout the International Human Microbiome Consortium, particularly the investigators of the MetaHIT project, for advancing human microbiome research. Data repository management was provided by the National Center for Biotechnology Information and the Intramural Research Program of the NIH National Library of Medicine. We appreciate the participation of the individuals from the Saint Louis, Missouri, and Houston, Texas areas who made this study possible. This research was supported in part by National Institutes of Health grants U54HG004969 to B.W.B.; U54HG003273 to R.A.G.; U54HG004973 to R.A.G., S.K.H. and J.F.P.; U54HG003067 to E.S.Lander; U54AI084844 to K.E.N.; N01AI30071 to R.L.Strausberg; U54HG004968 to G.M.W.; U01HG004866 to O.R.W.; U54HG003079 to R.K.W.; R01HG005969 to C.H.; R01HG004872 to R.K.; R01HG004885 to M.P.; R01HG005975 to P.D.S.; R01HG004908 to Y.Y.; R01HG004900 to M.K.Cho and P. Sankar; R01HG005171 to D.E.H.; R01HG004853 to A.L.M.; R01HG004856 to R.R.; R01HG004877 to R.R.S. and R.F.; R01HG005172 to P. Spicer.; R01HG004857 to M.P.; R01HG004906 to T.M.S.; R21HG005811 to E.A.V.; M.J.B. was supported by UH2AR057506; G.A.B. was supported by UH2AI083263 and UH3AI083263 (G.A.B., C. N. Cornelissen, L. K. Eaves and J. F. Strauss); S.M.H. was supported by UH3DK083993 (V. B. Young, E. B. Chang, F. Meyer, T. M. S., M. L. Sogin, J. M. Tiedje); K.P.R. was supported by UH2DK083990 (J. V.); J.A.S. and H.H.K. were supported by UH2AR057504 and UH3AR057504 (J.A.S.); DP2OD001500 to K.M.A.; N01HG62088 to the Coriell Institute for Medical Research; U01DE016937 to F.E.D.; S.K.H. was supported by RC1DE0202098 and R01DE021574 (S.K.H. and H. Li); J.I. was supported by R21CA139193 (J.I. and D. S. Michaud); K.P.L. was supported by P30DE020751 (D. J. Smith); Army Research Office grant W911NF-11-1-0473 to C.H.; National Science Foundation grants NSF DBI-1053486 to C.H. and NSF IIS-0812111 to M.P.; The Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231 for P.S. C.; LANL Laboratory-Directed Research and Development grant 20100034DR and the US Defense Threat Reduction Agency grants B104153I and B084531I to P.S.C.; Research Foundation - Flanders (FWO) grant to K.F. and J.Raes; R.K. is an HHMI Early Career Scientist; Gordon & Betty Moore Foundation funding and institutional funding from the J. David Gladstone Institutes to K.S.P.; A.M.S. was supported by fellowships provided by the Rackham Graduate School and the NIH Molecular Mechanisms in Microbial Pathogenesis Training Grant T32AI007528; a Crohn's and Colitis Foundation of Canada Grant in Aid of Research to E.A.V.; 2010 IBM Faculty Award to K.C.W.; analysis of the HMP data was performed using National Energy Research Scientific Computing resources, the BluBioU Computational Resource at Rice University.

Author Contributions Principal investigators: B.W.B., R.A.G., S.K.H., B.A.M., K.E.N., J.F.P., G.M.W., O.W., R.K.W. Manuscript preparation: D.G., C.H., R.K., O.W. Funding agency management: C.C.B., T.B., V.R.B., J.L.C., S.C., C.D., V.D.F., C.G., M.Y.G., R.D.L., J.M., P.M., J.P., L.M.P., J.A.S., L.W., C.W., K.A.W. Project leadership: S.A., J.H.B., B.W.B., A.T.C., H.H.C., A.M.E., M.G.F., R.S.F., D.G., M.G.G., K.H., S.K.H., C.H., E.A.L., R.M., V.M., J.C.M., B.A.M., M.M., D.M.M., K.E.N., J.F.P., E.J.S., J.V., G.M.W., O.W., A.M.W., K.C.W., J.R.W., S.K.Y., Q.Z. Analysis preparation for manuscript: J.C.C., K.F., D.G., A.G., K.H.H., C.H., R.K., D.K., H.H.K., O.K., K.P.L., R.E.L., J.R., J.F.S., P.D.S., N.S. Data release: L.A., T.B., I.A.C., K.C., H.H.C., N.J.D., D.J.D., A.M.E., V.M.F., L.F., J.M.G., S.G., S.K.H., M.E.H., C.J., V.J., C.K., A.A.M., V.M.M., T.M., M.M., D.M.M., J.O., K.P., J.F.P., C.P., X.Q., R.K.S., N.S., I.S., E.J.S., D.V.W., O.W., K.W., K.C.W., C.Y., B.P.Y., Q.Z. Methods and research development: S.A., H.M.A., M.B., D.M.C., A.M.E., R.L.E., M.F., S.F., M.G.F., D.C.F., D.G., G.G., B.J.H., S.K.H., M.E.H., W.A.K., N.L., K.L., V.M., E.R.M., B.A.M., M.M., D.M.M., C.N., J.F.P., M.E.P., X.Q., M.C.R., C.R., E.J.S., S.M.S., D.G.T., D.V.W., G.M.W., Y.W., K.M.W., S.Y., B.P.Y., S.K.Y., Q.Z. DNA sequence production: S.A., E.A., T.A., T.B., C.J.B., D.A.B., K.D.D., S.P.D., A.M.E., R.L.E., C.N.F., S.F., C.C.F., L.L.F., R.S.F., B.H., S.K.H., M.E.H., V.J., C.L.K., S.L.L., N.L., L.L., D.M.M., I.N., C.N., M.O., J.F.P., X.Q., J.G.R., Y.R., M.C.R., D.V.W., Y.W., B.P.Y., Y.Z. Clinical sample collection: K.M.A., M.A.C., W.M.D., L.L.F., N.G., H.A.H., E.L.H., J.A.K., W.A.K., T.M., A.L.M., P.M., S.M.P., J.F.P., G.A.S., J.V., M.A.W., G.M.W. Body site experts: K.M.A., E.A.V., G.A., L.B., M.J.B., C.C.D., F.E.D., L.F., J.I., J.A.K., S.K.H., H.H.K., K.P.L., P.J.M., J. Ravel, T.M.S., J.A.S., J.D.S., J.V. Ethical, legal and social implications: R.M.F., D.E.H., W.A.K., N.B.K., C.M.L., A.L.M., R.R., P. Sankar, R.R.S., P. Spicer, L.Z. Strain management: E.A.V., J.H.B., I.A.C., K.C., S.W.C., H.H.C., T.Z.D., A.S.D., A.M.E., M.G.F., M.G.G., S.K.H., V.J., N.C.K., S.L.L., L.L., K.L., E.A.L., V.M.M., B.A.M., D.M.M., K.E.N., I.N., I.P., L.S., E.J.S., C.M.T., M.T., D.V.W., G.M.W., A.M.W., Y.W., K.M.W., B.P.Y., L.Z., Y.Z. 16S data analysis: K.M.A., E.J.A., G.L.A., C.A.A., M.B., B.W.B., J.P.B., G.A.B., S.R.C., S.C., J.C., T.Z.D., F.E.D., E.D., A.M.E., R.C.E., K.F., M.F., A.A.F., J.F., H.G., D.G., B.J.H., T.A.H., S.M.H., C.H., J.I., J.K.J., S.T.K., S.K.H., R.K., H.H.K., O.K., P.S.L., R.E.L., K.L., C.A.L., D.M., B.A.M., K.A.M., M.M., M.P., J.F.P., M.P., K.S.P., X.Q., J. Raes, K.P.R., M.C.R., B.R., J.F.S., P.D.S., T.M.S., N.S., J.A.S., W.D.S., T.J.S., C.S.S., E.J.S., R.M.T., J.V., T.A.V., Z.W., D.V.W., G.M.W., J.R.W., K.M.W., Y.Y., S.Y., Y.Z. Shotgun data processing and alignments: C.J.B., J.C.C., E.D., D.G., A.G., M.E.H., H.J., D.K., K.C.K., C.L.K., Y.L., J.C.M., B.A.M., M.M., D.M.M., J.O., J.F.P., X.Q., J.G.R., R.K.S., N.U.S., I.S., E.J.S., G.G.S., S.M.S., J.W., Z.W., G.M.W., O.W., K.C.W., T.W., S.K.Y., L.Z. Assembly: H.M.A., C.J.B., P.S.C., L.C., Y.D., S.P.D., M.G.F., M.E.H., H.J., S.K., B.L., Y.L., C.L., J.C.M., J.M.M., J.R.M., P.J.M., M.M., J.F.P., M.P., M.E.P., X.Q., M.R., R.K.S., M.S., D.D.S., G.G.S., S.M.S., C.M.T., T.J.T., W.W., G.M.W., K.C.W., L.Y., Y.Y., S.K.Y., L.Z. Annotation: O.O.A., V.B., C.J.B., I.A.C., A.T.C., K.C., H.H.C., A.S.D., M.G.G., J.M.G., J.G., A.G., S.G., B.J.H., K.H., S.K.H., C.H., H.J., N.C.K., R.M., V.M.M., K.M., T.M., M.M., J.O., K.P., M.P., X.Q., N.S., E.J.S., G.G.S., S.M.S., M.T., G.M.W., K.C.W., J.R.W., C.Y., S.K.Y., Q.Z., L.Z., W.G.S. Metabolic reconstruction: S.A., B.L.C., J.G., C.H., J.I., B.A.M., M.M., B.R., A.M.S., N.S., M.T., G.M.W., S.Y., Q.Z., J.D.Z.

Author Information All data used in this study is available from the Human Microbiome Project Data Analysis and Coordination Center at http://hmpdacc.org and from the NCBI. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.H. (chuttenh@hsph.harvard.edu).

The Human Microbiome Project Consortium

Curtis Huttenhower[1,2]*, Dirk Gevers[2]*, Rob Knight[3,4], Sahar Abubucker[5], Jonathan H. Badger[6], Asif T. Chinwalla[5], Heather H. Creasy[7], Ashlee M. Earl[2], Michael G. FitzGerald[2], Robert S. Fulton[5], Michelle G. Giglio[7], Kymberlie Hallsworth-Pepin[5], Elizabeth A. Lobos[5], Ramana Madupu[6], Vincent Magrini[5], John C. Martin[5], Makedonka Mitreva[5], Donna M. Muzny[8], Erica J. Sodergren[5], James Versalovic[9,10], Aye M. Wollam[5], Kim C. Worley[8], Jennifer R. Wortman[7], Sarah K. Young[2], Qiandong Zeng[2], Kjersti M. Aagaard[11], Olukemi O. Abolude[7], Emma Allen-Vercoe[12], Eric J. Alm[13,2], Lucia Alvarado[2], Gary L. Andersen[14], Scott Anderson[2], Elizabeth Appelbaum[5], Harindra M. Arachchi[2], Gary Armitage[15], Cesar A. Arze[7], Tulin Ayvaz[16], Carl C. Baker[17], Lisa Begg[18], Tsegahiwot Belachew[19], Veena Bhonagiri[5], Monika Bihan[6], Martin J. Blaser[20], Toby Bloom[2], Vivien Bonazzi[21], J. Paul Brooks[22,23], Gregory A. Buck[23,24], Christian J. Buhay[8], Dana A. Busam[6], Joseph L. Campbell[21,19], Shane R. Canon[25], Brandi L. Cantarel[7], Patrick S. G. Chain[26,27], I-Min A. Chen[28], Lei Chen[5], Shaila Chhibba[21], Ken Chu[28], Dawn M. Ciulla[2], Jose C. Clemente[3], Sandra W. Clifton[5], Sean Conlan[79], Jonathan Crabtree[7], Mary A. Cutting[29], Noam J. Davidovics[7], Catherine C. Davis[30], Todd Z. DeSantis[31], Carolyn Deal[19], Kimberley D. Delehaunty[5], Floyd E. Dewhirst[32,33], Elena Deych[34], Yan Ding[8], David J. Dooling[5], Shannon P. Dugan[8], Wm Michael Dunne[35,36], A. Scott Durkin[6], Robert C. Edgar[37], Rachel L. Erlich[2], Candace N. Farmer[5], Ruth M. Farrell[38], Karoline Faust[39,40], Michael Feldgarden[2], Victor M. Felix[7], Sheila Fisher[2], Anthony A. Fodor[41], Larry J. Forney[42], Leslie Foster[6], Valentina Di Francesco[19], Jonathan Friedman[43], Dennis C. Friedrich[2], Catrina C. Fronick[5], Lucinda L. Fulton[5], Hongyu Gao[5], Nathalia Garcia[44], Georgia Giannoukos[2], Christina Giblin[19], Maria Y. Giovanni[19], Jonathan M. Goldberg[2], Johannes Goll[6], Antonio Gonzalez[45], Allison Griggs[2], Sharvari Gujja[2], Susan Kinder Haake[46], Brian J. Haas[2], Holli A. Hamilton[29], Emily L. Harris[29], Theresa A. Hepburn[2], Brandi Herter[5], Diane E. Hoffmann[47], Michael E. Holder[8], Clinton Howarth[2], Katherine H. Huang[2], Susan M. Huse[48], Jacques Izard[32,33], Janet K. Jansson[49], Huaiyang Jiang[8], Catherine Jordan[7], Vandita Joshi[8], James A. Katancik[50], Wendy A. Keitel[16], Scott T. Kelley[51], Cristyn Kells[2], Nicholas B. King[52], Dan Knights[45], Heidi H. Kong[53], Omry Koren[54], Sergey Koren[55], Karthik C. Kota[5], Christie L. Kovar[8], Nikos C. Kyrpides[27], Patricio S. La Rosa[34], Sandra L. Lee[8], Katherine P. Lemon[32,56], Niall Lennon[2], Cecil M. Lewis[57], Lora Lewis[8], Ruth E. Ley[54], Kelvin Li[6], Konstantinos Liolios[27], Bo Liu[55], Yue Liu[8], Chien-Chi Lo[26], Catherine A. Lozupone[3], R. Dwayne Lunsford[29], Tessa Madden[58], Anup A. Mahurkar[7], Peter J. Mannon[59], Elaine R. Mardis[5], Victor M. Markowitz[27,28], Konstantinos Mavromatis[27], Jamison M. McCorrison[6], Daniel McDonald[3], Jean McEwen[21], Amy L. McGuire[60], Pamela McInnes[29], Teena Mehta[2], Kathie A. Mihindukulasuriya[5], Jason R. Miller[6], Patrick J. Minx[5], Irene Newsham[8], Chad Nusbaum[2], Michelle O'Laughlin[5], Joshua Orvis[7], Ioanna Pagani[27], Krishna Palaniappan[28], Shital M. Patel[61], Matthew Pearson[2], Jane Peterson[21], Mircea Podar[62], Craig Pohl[5], Katherine S. Pollard[63,64,65], Mihai Pop[55,66], Margaret E. Priest[2], Lita M. Proctor[21], Xiang Qin[8], Jeroen Raes[39,40], Jacques Ravel[7], Jeffrey G. Reid[8], Mina Rho[67], Rosamond Rhodes[68], Kevin P. Riehle[69], Maria C. Rivera[23,24], Beltran Rodriguez-Mueller[51], Yu-Hui Rogers[6], Matthew C. Ross[16], Carsten Russ[2], Ravi K. Sanka[6], Pamela Sankar[70], J. Fah Sathirapongsasuti[1], Jeffery A. Schloss[21], Patrick D. Schloss[71], Thomas M. Schmidt[72], Matthew Scholz[26], Lynn Schriml[7], Alyxandria M. Schubert[71], Nicola Segata[1], Julia A. Segre[79], William D. Shannon[34], Richard R. Sharp[38], Thomas J. Sharpton[63], Narmada Shenoy[2], Nihar U. Sheth[23], Gina A. Simone[73], Indresh Singh[6], Christopher S. Smillie[43], Jack D. Sobel[74], Daniel D. Sommer[55], Paul Spicer[57], Granger G. Sutton[6], Sean M. Sykes[2], Diana G. Tabbaa[2], Mathangi Thiagarajan[6], Chad M. Tomlinson[5], Manolito Torralba[6], Todd J. Treangen[75], Rebecca M. Truty[63], Tatiana A. Vishnivetskaya[62], Jason Walker[5], Lu Wang[21], Zhengyuan Wang[5], Doyle V. Ward[2], Wesley Warren[5], Mark A. Watson[35], Christopher Wellington[21], Kris A. Wetterstrand[21], James R. White[7], Katarzyna Wilczek-Boney[8], YuanQing Wu[8], Kristine M. Wylie[5], Todd Wylie[5], Chandri Yandava[2], Liang Ye[5], Yuzhen Ye[67], Shibu Yooseph[76], Bonnie P. Youmans[16], Lan Zhang[8], Yanjiao Zhou[5], Yiming Zhu[8], Laurie Zoloth[77], Jeremy D. Zucker[2], Bruce W. Birren[2], Richard A. Gibbs[8], Sarah K. Highlander[8,16], Barbara A. Methé[6], Karen E. Nelson[6], Joseph F. Petrosino[8,78,16], George M. Weinstock[5], Richard K. Wilson[5] & Owen White[7]

[1]Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [2]The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA. [4]Howard Hughes Medical Institute, Boulder, Colorado 80309, USA. [5]The Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [6]J. Craig Venter Institute, Rockville, Maryland 20850, USA. [7]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [8]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [9]Department of Pathology & Immunology, Baylor College of Medicine, Houston, Texas 77030, USA. [10]Department of Pathology, Texas Children's Hospital, Houston, Texas 77030, USA. [11]Department of Obstetrics & Gynecology, Division of Maternal-Fetal Medicine, Baylor College of Medicine, Houston, Texas 77030, USA. [12]Molecular and Cellular Biology, University of Guelph, Guleph, Ontario N1G 2W1, Canada. [13]Department of Civil & Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [14]Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [15]School of Dentistry, University of California, San Francisco, San Francisco, California 94143, USA. [16]Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas 77030, USA. [17]National Institute of Arthritis and Musculoskeletal and Skin, National Institutes of Health, Bethesda, Maryland 20892, USA. [18]Office of Research on Women's Health, National Institutes of Health, Bethesda, Maryland 20892, USA. [19]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. [20]Department of Medicine, New York University Langone Medical Center, New York, New York 10016, USA. [21]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [22]Department of Statistical Sciences and Operations Research, Virginia Commonwealth University, Richmond, Virginia 23284, USA. [23]Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23284, USA. [24]Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284, USA. [25]Technology Integration Group, National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [26]Genome Science Group, Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. [27]Joint Genome Institute, Walnut Creek, California 94598, USA. [28]Biological Data Management and Technology Center, Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [29]National Institute of Dental and Craniofacial Research (NIDCR), National Institutes of Health, Bethesda, Maryland 20892, USA. [30]FemCare Product Safety and Regulatory Affairs, The Procter & Gamble Company, Cincinnati, Ohio 45224, USA. [31]Bioinformatics Department, Second Genome, Inc., San Bruno, California 94066, USA. [32]Department of Molecular Genetics, Forsyth Institute, Cambridge, Massachusetts 02142, USA. [33]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, Massachusetts 02115, USA. [34]Department of Medicine, Division of General Medical Science, Washington University School of Medicine, St. Louis, Missouri 63110, USA. [35]Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, Missouri 63110, USA. [36]bioMerieux, Inc., Durham, South Carolina 27712, USA. [37]drive5.com, Tiburon, California 94920, USA. [38]Center for Ethics, Humanities and Spiritual Care, Cleveland Clinic, Cleveland, Ohio 44195, USA. [39]Department of Structural Biology, VIB, Belgium, 1050 Ixelles, Belgium. [40]Department of Applied Biological Sciences (DBIT), Vrije Universiteit Brussel, 1050 Ixelles, Belgium. [41]Department of Bioinformatics and Genomics, University of North Carolina - Charlotte, Charlotte, North Carolina 28223, USA. [42]Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844, USA. [43]Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [44]Center for Advanced Dental Education, Saint Louis University, St. Louis, Missouri 63104, USA. [45]Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA. [46]Division of Associated Clinical Specialties and Dental Research Institute, UCLA School of Dentistry, Los Angeles, California 90095, USA. [47]University of Maryland Francis King Carey School of Law, Baltimore, Maryland 21201, USA. [48]Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA. [49]Ecology Department, Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [50]Department of Periodontics, University of Texas Health Science Center School of Dentistry, Houston, Texas 77030, USA. [51]Department of Biology, San Diego State University, San Diego, California 92182, USA. [52]Faculty of Medicine, McGill University, 3647 Peel St, Montreal, Quebec H3A 1X1, Canada. [53]Dermatology Branch, CCR, National Cancer Institute, Bethesda, Maryland 20892, USA. [54]Department of Microbiology, Cornell University, Ithaca, New York 14853, USA. [55]Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland 20742, USA. [56]Division of Infectious Diseases, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts 02115, USA. [57]Department of Anthropology, University of Oklahoma, Norman, Oklahoma 73019, USA. [58]Department of Obstetrics and Gynecology, Washington University School of Medicine, Saint Louis, Missouri 63110, USA. [59]Division of Gastroenterology and Hepatology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [60]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas 77030, USA. [61]Medicine-Infectious Disease, Baylor College of Medicine, Houston, Texas 77030, USA. [62]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA. [63]Gladstone Institutes, University of California, San Francisco, San Francisco, California 94158, USA. [64]Institute for Human Genetics, University of California, San Francisco, San Francisco, California 94158, USA. [65]Division of Biostatistics, University of California, San Francisco, San Francisco, California 94158, USA. [66]Department of Computer Science, University of Maryland, College Park, Maryland 20742, USA. [67]School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA. [68]Mount Sinai School of Medicine, New York, New York 10029, USA. [69]Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. [70]Center for Bioethics and Department of Medical Ethics, University of Pennsylvania, Philadelphia, Pennsylviana 19104, USA. [71]Department of Microbiology & Immunology, University of Michigan, Ann Arbor, Michigan 48109, USA. [72]Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824, USA. [73]The EMMES Corporation, Rockville, Maryland 20850, USA. [74]Harper University Hospital, Wayne State University School of Medicine, Detroit, Michigan 48201, USA. [75]McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [76]J. Craig Venter Institute, San Diego, California 92121, USA. [77]Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA. [78]Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, Texas 77030, USA. [79]Genetics and Molecular Biology Branch, National Human Genome Research Institute, Bethesda, Maryland 20892, USA.
*These authors contributed equally to this work.

ARTICLE

doi:10.1038/nature11209

A framework for human microbiome research

The Human Microbiome Project Consortium*

A variety of microbial communities and their genes (the microbiome) exist throughout the human body, with fundamental roles in human health and disease. The National Institutes of Health (NIH)-funded Human Microbiome Project Consortium has established a population-scale framework to develop metagenomic protocols, resulting in a broad range of quality-controlled resources and data including standardized methods for creating, processing and interpreting distinct types of high-throughput metagenomic data available to the scientific community. Here we present resources from a population of 242 healthy adults sampled at 15 or 18 body sites up to three times, which have generated 5,177 microbial taxonomic profiles from 16S ribosomal RNA genes and over 3.5 terabases of metagenomic sequence so far. In parallel, approximately 800 reference strains isolated from the human body have been sequenced. Collectively, these data represent the largest resource describing the abundance and variety of the human microbiome, while providing a framework for current and future studies.

Advances in sequencing technologies coupled with new bioinformatic developments have allowed the scientific community to begin to investigate the microbes that inhabit our oceans, soils, the human body and elsewhere[1]. Microbes associated with the human body include eukaryotes, archaea, bacteria and viruses, with bacteria alone estimated to outnumber human cells within an individual by an order of magnitude. Our knowledge of these communities and their gene content, referred to collectively as the human microbiome, has until now been limited by a lack of population-scale data detailing their composition and function.

The US NIH-funded Human Microbiome Project Consortium (HMP) brought together a broad collection of scientific experts to explore these microbial communities and their relationships with their human hosts. As such, the HMP[2] has focused on producing reference genomes (viral, bacterial and eukaryotic), which provide a critical framework for subsequent metagenomic annotation and analysis, and on generating a baseline of microbial community structure and function from an adult cohort defined by a carefully delineated set of clinical inclusion and exclusion criteria that we term 'healthy' in this study (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd002854.2). Investigations of the microbiome from this cohort incorporated several complementary analyses including: 16S ribosomal RNA (rRNA) gene sequence (16S) and taxonomic profiles, whole-genome shotgun (WGS) or metagenomic sequencing of whole community DNA, and alignment of the assembled sequences to the reference microbial genomes from the human body[3,4]. Thus, the HMP complements other large-scale sequence-based human microbiome projects such as the MetaHIT project[5], which focused on examination of the gut microbiome using WGS data including samples from cohorts exhibiting a wide range of health statuses and physiological characteristics.

Additional projects supported by the HMP are investigating the association of specific components and dynamics of the microbiome with a variety of disease conditions, developing tools and technology including isolating and sequencing uncultured organisms, and studying the ethical, legal and social implications of human microbiome research (http://commonfund.nih.gov/hmp/fundedresearch.aspx). A comprehensive list of current publications from HMP projects is available at http://commonfund.nih.gov/hmp/publications.aspx.

Here we detail the resources created so far by the HMP initiative including: clinical specimens (samples), reference genomes, sequencing and annotation protocols, methods and analyses. We describe the thousands of samples obtained from 15 or 18 distinct body sites from 242 donors over multiple time points that were processed at two clinical centres (Baylor College of Medicine (BCM) and Washington University School of Medicine). We also describe the laboratory and computational protocols developed for reliably generating and interpreting the human microbiome data. HMP resources include both protocols for, and the subsequent data generated from, 16S and metagenomic sequencing of human microbiome samples. During this study, these protocols were rigorously standardized and quality controlled for simultaneous use across four sequencing centres (BCM Human Genome Sequencing Center, The Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, the J. Craig Venter Institute and The Genome Institute at Washington University School of Medicine). In particular, we focus on the production of the first phase of metagenomic data sets (phase I) used for subsequent in-depth analyses, and we summarize standards and recommendations based on our experiences generating and analysing these data. An additional set of publications (many included in the references and in those of ref. 4) describe in further detail the microbial ecology and microbiological implications of these data. Collectively these resources and analyses represent an important framework for human microbiome research.

HMP resource organization

Supplementary Fig. 1 summarizes organization of the HMP, including the data processing and analytical steps, and the scientific entities gathered to conduct the project. An overview of available HMP data sets and additional resources are provided in Supplementary Tables 1–3. Donors were recruited and enrolled into the HMP through the two clinical centres. Over 240 adults were carefully screened and phenotyped before sampling one to three times at 15 (male) or 18 (female) body sites using a common sampling protocol (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd003190.2). All included subjects were between the ages of 18 and 40 years and had passed a

*Lists of participants and their affiliations appear at the end of the paper.

screening for systemic health based on oral, cutaneous and body mass exclusion criteria (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/GetPdf.cgi?id=phd002854.2) (K. Aagaard *et al.*, manuscript submitted).

A Data Analysis and Coordination Center (DACC) was created to serve as the central repository for all HMP WGS, 16S and reference genome sequence information generated by the four sequencing centres. The DACC supports access to analysis software, biological samples, clinical protocols, news, publication announcements and project statistics, and performed centralized analysis of HMP reference genome and WGS annotation in cooperation with the sequencing centres. All unprocessed 16S, WGS and reference genome sequence data are deposited at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/bioproject/43021). Unless otherwise noted, all data sets and protocols described here are available to the scientific community at the DACC (http://hmpdacc.org). Specific data sets referred to in this work and available at the DACC are indicated in parentheses with the preface 'RES'.

Phase I 16S and WGS sequencing overview

A set of 5,298 samples were collected from 242 adults (K. Aagaard *et al.*, manuscript submitted; Table 1 and Supplementary Table 4), from which 16S and WGS data were generated for a total of 5,177 taxonomically characterized communities (16S) and 681 WGS samples describing the microbial communities from habitats within the human airways, skin, oral cavity, gut and vagina. For a subset of 560 samples, both data types were generated (Table 1). These efforts constitute our initial primary metagenomic data sets (phase I) described in more detail later. Additional efforts are ongoing to sequence and analyse the remaining samples from the complete HMP collection (11,174 primary specimens in total from 300 individuals sampled up to three times over 22 months) (K. Aagaard *et al.*, manuscript submitted).

16S standards development and sequencing

The goals of the HMP required that 16S sequences and profiles from data produced at the four participating sequencing centres be comparable in a variety of downstream analyses; however, no suitable methodology was available at the commencement of the project. While establishing 16S protocols, we determined that many components of data production and processing can contribute errors and artefacts. We investigated methods that avoid these errors and their subsequent effects on taxonomic classification and operational taxonomic unit (OTU)-based community structure. The results are discussed in detail in Supplementary Information and ref. 6. Thus, multiple evaluations of 16S protocols were undertaken before adopting a single standardized protocol that ensured consistency in the high-throughput production.

To maximize accuracy and consistency, protocols were evaluated primarily using a synthetic mock community of 21 known organisms[6] (Supplementary Table 5). Additional testing of the protocol was carried out on a subset of HMP samples (Supplementary Table 1). Collectively, these efforts resulted in adoption of a protocol to amplify and sequence samples using the Roche-454 FLX Titanium platform[6] (http://www.hmpdacc.org/doc/HMP_MDG_454_16S_Protocol.pdf). The HMP created both cell mixtures and genomic DNA extracts of the mock community (Supplementary Tables 2 and 5). A large body of metagenomic data (both 16S and WGS) (RES:HMMC) from these and other calibration experiments are available to the community to facilitate further benchmarking of new molecular and analytical approaches (Supplementary Table 3).

The majority of the sample collection was targeted for 16S sequencing using the 454 FLX Titanium based strategy[6]. The nucleotide sequence of the 16S rRNA gene consists of regions of highly conserved sequence, which alternate with nine regions or windows of variable nucleotide sequence that constitute the most informative portions of the gene sequence for use in taxonomic classification. A window covering number three (V3) to five (V5) variable regions (V35) of the 16S rRNA gene was chosen as the target for 4,879 samples. Sequence of a V1 to V3 (V13) window was also included for a subset of 2,971 samples to provide a complementary view of taxonomic profiles[6] (RES:HMR16S) (Table 1, Supplementary Figs 2, 3 and Supplementary Information).

After adoption of the 16S protocol, including removal of multiple sources of potential artefacts or bias generated by 16S sequencing using pyrosequencing[7,8], a variety of approaches for accurate diversity estimation were developed and compared[9]. A 16S data processing pipeline was established using the mothur software package[10] (Supplementary Information), which includes two optional low and high stringency approaches. The former provides an output favouring longer read lengths tailored towards taxonomic classification, the latter an output with more aggressive sequence error reduction tailored towards OTU construction (RES:HMMCP). A third complementary pipeline was also developed using the QIIME software package[11] (Supplementary Information), which processes these data using an

Table 1 | HMP donor samples examined by 16S and WGS

Body region	Body site	Total samples	Total 16S samples	V13 samples	V13 read depth (M)*	V35 samples	V35 read depth (M)*	Samples V13 and V35	Total WGS samples	Total read depth (G)†	Filtered reads (%)‡	Human reads (%)§	Remaining read depth (G)†	Samples 16S and WGS
Gut	Stool	352	337	193	1.4	328	2.4	184	136	1,720.7	15	1	1,450.6	124
Oral cavity	Buccal mucosa	346	330	184	1.3	314	1.7	168	107	1,438.0	9	82	136.7	91
	Hard palate	325	325	179	1.2	310	1.7	164	1	10.9	20	25	5.9	1
	Keratinized gingiva	335	329	183	1.3	319	1.7	173	6	72.3	5	47	34.4	0
	Palatine tonsils	337	332	189	1.2	315	1.9	172	6	74.8	2	80	13.5	1
	Saliva	315	310	166	0.9	292	1.5	148	5	55.7	1	91	4.2	0
	Subgingival plaque	334	328	186	1.2	314	1.8	172	7	92.1	5	79	15.3	1
	Supragingival plaque	345	331	192	1.3	316	1.9	177	115	1,500.7	15	40	674.8	101
	Throat	331	325	176	1.0	312	1.7	163	7	78.8	4	79	13.6	1
	Tongue dorsum	348	332	193	1.3	320	2.0	181	122	1,620.1	15	19	1,084.3	106
Airway	Anterior nares	316	302	169	1.0	283	1.2	150	84	1,129.9	3	96	14.3	70
Skin	Left antecubital fossa	269	269	158	0.7	221	0.5	110	0	NA	NA	NA	0	NA
	Left retroauricular crease	313	312	188	1.6	295	1.5	171	9	126.3	9	73	22.1	8
	Right antecubital fossa	274	274	158	0.7	229	0.5	113	0	NA	NA	NA	0	NA
	Right retroauricular crease	319	316	190	1.4	304	1.6	178	15	181.9	18	59	42.4	12
Vagina	Mid-vagina	145	143	91	0.6	140	1.0	88	2	22.6	0	99	0.2	0
	Posterior fornix	152	142	89	0.6	136	1.0	83	53	702.1	6	90	25.2	43
	Vaginal introitus	142	140	87	0.6	131	0.9	78	3	36.5	1	98	0.6	1
	Total	5,298	5,177	2,971	19	4,879	26.3	2,673	681	8,863.3	11	49	3,538.1	560

NA, not applicable.
* 1×10^6 reads post-processing with the mothur pipeline (Supplementary Information).
† 1×10^9 reads (Supplementary Information).
‡ Fraction of reads with low quality bases that were removed (Supplementary Information).
§ Fraction of human reads that were removed (Supplementary Information).

OTU-binning strategy to which taxonomic classification is added (RES:HMQCP). All pipelines result in highly comparable views of the human microbiome.

Metagenomic assembly and gene cataloguing

Approximately 749 samples representing targeted body sites were chosen for WGS sequencing using the Illumina GAIIx platform with 101-base-pair paired-end reads. From a high-quality set of 681 samples an average depth of 13 Gb (± 4.3) was achieved per sample, collectively producing a total of 8.8 Tb (RES:HMIWGS) (Table 1). Theoretically, these per sample data are sufficient to cover a 3 Mb bacterial genome present at only 0.8% abundance with a probability of 90% (M. C. Wendl *et al.*, manuscript submitted). In addition, 12 stool samples were simultaneously sequenced using the 454 FLX Titanium platform (RES:HM4WGS). Comparisons between the centres demonstrated high consistency of target sequencing depth and success rates[4]. After development of a protocol for removing reads resulting from human DNA contamination (Supplementary Information), 49% of the reads were targeted for removal as human (for information on authorized access to these reads, see Supplementary Information). Samples collected from soft tissue tended to have higher human contamination (for example, mid-vagina (96%), anterior nares (82%) and throat (75%)). Preparations from saliva were also high in human DNA sequence (80%), whereas stool contained a relatively low abundance of human reads (up to 1%) (Supplementary Fig. 4).

After application of a quality control protocol that includes human sequence removal, quality filtering and trimming of reads (Supplementary Information), the remaining 3.5 Tb from 681 samples were subjected to a three-tiered complementary analysis strategy (Supplementary Information) of reference genome mapping (which was able to use ~57% of the data), assembly and gene prediction (~50% of the data), and metabolic reconstruction (~36% of the data). This combined strategy facilitated the extraction of maximal organismal and functional information.

Metagenomic assemblies were generated for all available samples using an optimized SOAPdenovo protocol with parameters designed to produce substrates for downstream analyses such as gene and function prediction, resulting in a total of 41 million contigs (RES:HMASM) (Supplementary Information). Reads that remained unassembled were pooled across individual body sites and re-assembled using the same approach, resulting in an additional 4,200,672 contigs (RES:HMBSA). These body-site-specific assemblies are aimed at reconstructing organisms that represent too small a fraction in any individual sample to assemble but are found among many individuals. For 12 stool samples both Illumina and 454 FLX Titanium data (RES:HM4WGS) were generated, allowing a hybrid assembly approach using Newbler (Supplementary Information) (RES:HMHASM). Overall, the assembly statistics recovered varied substantially depending on body site and community complexity (Supplementary Fig. 5). However, our results indicate that, for the assembly strategy we used, metagenomic assembly quality plateaus at approximately 6 Gb of microbial sequence coverage for a sample possessing a microbial community structure similar to that of stool samples (Supplementary Fig. 6).

A WGS-based perspective of community membership was obtained by aligning the reads to a set of 1,742 finished bacterial, 131 archaeal, 3,683 viral and 326 microeukaryotic reference genomes[12] (RES:HMREFG) (Supplementary Information) representing a broad taxonomic range from each of these four domains. A total of 57.6% of the high-quality microbial reads could be associated with a known genome (ranging from 33–77% for anterior nares and posterior fornix, respectively) (RES:HMSCP). The overwhelming majority of mapped sequences originated from bacteria (99.7%), while the remaining reads mapped to microeukaryotes (0.3%) or archaea (<0.01%) (Supplementary Information).

Two complementary approaches were used to summarize overall function and metabolism of the human microbiome, producing two primary data sets of annotations (RES:HMMRC and RES:HMGI) (Supplementary Information) and additional secondary analyses (RES:HMGS, HMHGI, HMGC and HMGOI) (Supplementary Information) available to the community for further interrogation. The first primary data set of annotations was produced by mapping individual shotgun reads to characterized protein families[13] (RES:HMMRC). The second was produced from functionally annotated gene predictions generated from the metagenomic assemblies (RES:HMGI), which were subsequently grouped according to high-level biological processes and to selected additional processes specific to metabolism and regulation[14] (RES:HMGS) (Supplementary Tables 6, 7 and Supplementary Fig. 7).

HMP data generation and analysis lessons

A key manner in which the HMP resources will serve to guide future studies of the microbiome is by enabling informed decisions regarding sampling protocols and genomic DNA preparation (K. Aagaard *et al.*, manuscript submitted), sequencing depth (M. C. Wendl *et al.*, manuscript submitted), statistical power (P. S. La Rosa *et al.*, manuscript submitted) and metagenomic data type. As indicated in Table 1, the consortium successfully amplified 16S sequences to our target depth at all 18 body sites, with the fewest sequences recovered consistently from the antecubital fossae. The amount of host human DNA recovered and the finest level of OTU resolution varied for 16S sequences among body sites[6] (Supplementary Figs 3 and 4).

From our WGS investigations, a series of protocols (http://hmpdacc.org/tools_protocols/tools_protocols.php) have been established to process large volumes of short-read WGS data and to annotate and examine these data through both a multi-tiered assembly approach and as single reads[15]. An investigator's choice of metagenomic technologies can thus be guided not only by a 16S versus WGS dichotomy, but also by the expected fraction of host sequence and the appropriate 16S region targeting the dominant taxa at each body site (Supplementary Figs 2–6 and 8).

Together, these data sets represent comprehensive and complementary views of the human microbiome, as shown by comparing organismal (Fig. 1a) and gene (Fig. 1b) catalogues, and the ratio of genes contributed per OTU (Fig. 1c). The discovery rate of new gene clusters (as determined by annotation of assembled WGS data) is in general detected more slowly relative to organismal discovery (as determined by OTU data) owing to the fragmentary nature of these community reads and assemblies despite high sequence depth (Fig. 1a, b and Supplementary Fig. 9), and the number of genes contributed per OTU varies by body site (Fig. 1c and Supplementary Information). However, in general, these results highlight an important point for consideration of further microbiome investigations using these data sets, as they suggest that the majority of the common taxa and genes present in this reference population have been detected.

We additionally compared the gut community gene catalogue sampled by the HMP with that of MetaHIT in terms of total detected gene counts. The HMP recovered more total non-redundant gene counts (5,140,472) than reported by MetaHIT (3,299,822)[5], probably reflecting a combination of the increased sequence depth obtained by the HMP (11.7 Gb HMP, 4.5 Gb MetaHIT on average) and differences in data generation and processing[5].

The two non-redundant sets of gene sequences were subsequently combined and compared by matches to a database of orthologous groups[16] of functionally annotated genes. Approximately 57% of the orthologous groups recovered by this method overlapped between the data sets, while an additional 34% versus 10% were unique to the HMP and MetaHIT, respectively (Supplementary Fig. 10, Supplementary Table 8 and Supplementary Information). After removal of genes that received any orthologous group assignment, the remaining novel

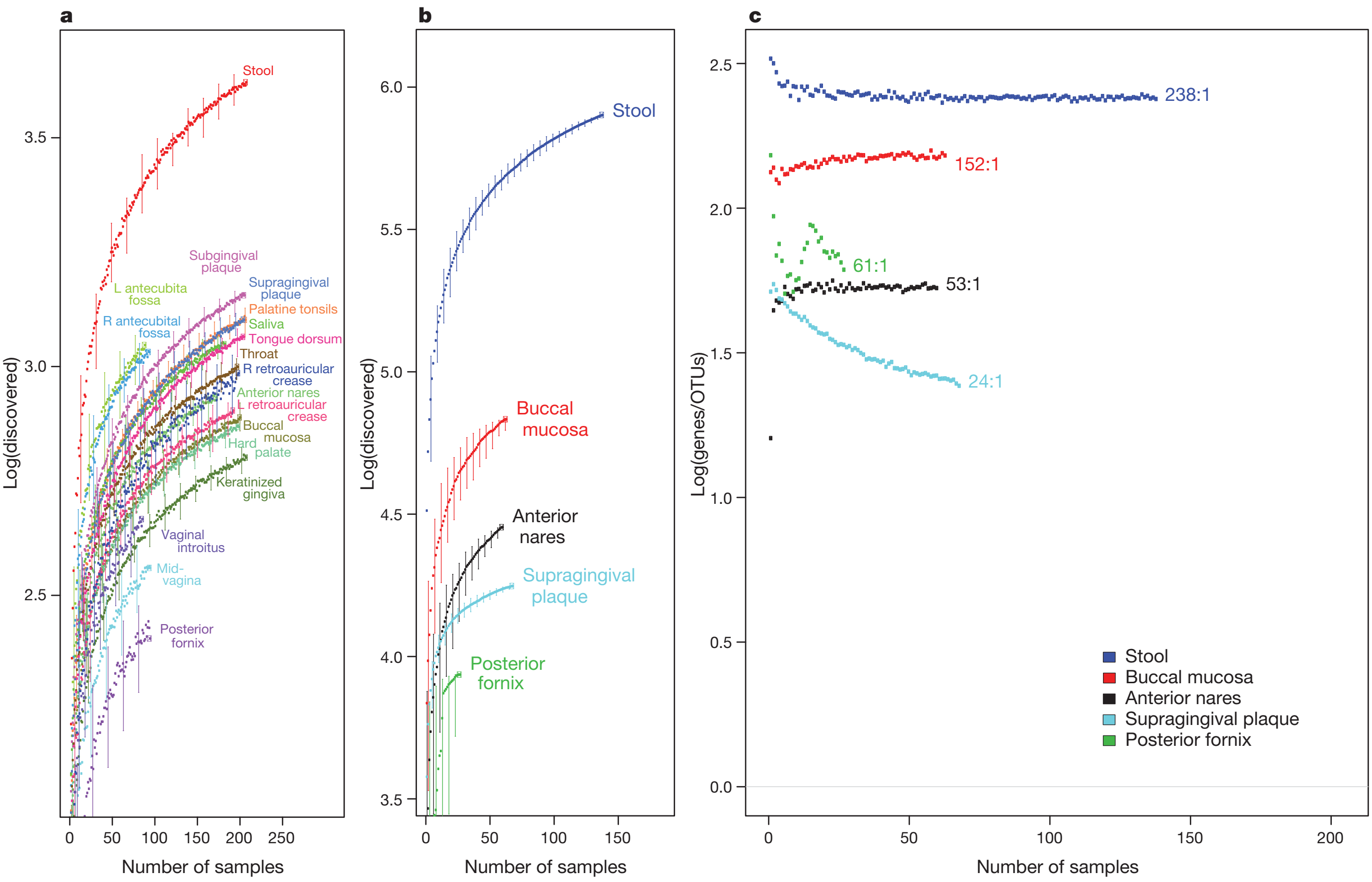


Figure 1 | Rates of gene and OTU discovery from HMP taxonomic and metagenomic data. a–c, Accumulation curves for OTU counts from 16S data (all body sites) (**a**), clustered gene index counts from metagenomic data (all applicable body sites) (**b**) and the ratio of average unique genes contributed versus unique OTUs encountered with increasing sample counts (**c**) (Supplementary Information). L, left; R, right. Ratios given for each curve in **c** represent the average number of unique genes contributed per unique OTU at the final sample count. Curves for stool, buccal mucosa and anterior nares suggest that the proportion of gene-to-taxa discovery has stabilized. In contrast, the curve for supragingival plaque suggests that relatively fewer new genes are being contributed per additional OTU. Error bars represent 95% confidence intervals.

genes were subsequently clustered[17]. Approximately 79% of the HMP-derived novel gene clusters were orthologous to one or more clusters in MetaHIT, while an additional 16% were unique to this study versus 5% for MetaHIT-derived data[5] (Supplementary Fig. 11, Supplementary Table 8 and Supplementary Information). These results suggest that, for this body habitat, relatively similar gene catalogues were recovered despite differences in experimental design and protocols. However, a greater proportion of both annotated and unique novel genes were detected in the HMP data set, emphasizing the utility of sequencing depth in recovering gene function and, in particular, deriving rare function. These results further underscore the importance of large-scale sequence-based studies of the microbiome to characterize better its gene content and diversity.

Human microbiome reference genomes

The current goal for the reference genome component of the HMP is to sequence at least 3,000 reference bacterial genomes, and additional viral and microeukaryotic genomes, associated with the human body. Thus far, more than 800 genomes have been sequenced and are available from the NCBI and the DACC (http://hmpdacc.org/HMRGD). From an alignment of WGS reads to reference genomes (RES:HMREFG), approximately 26% from the total read set (46% of all reads that could be aligned) were matched to a subset of 223 HMP reference genomes (Supplementary Information and Supplementary Data).

We continue to solicit community feedback for strains that will best benefit our attempts at understanding the breadth of human microbiome diversity. For example, a prioritized list of the 'most wanted' HMP taxa is being maintained (http://hmpdacc.org/most_wanted/) with the goal of targeting these difficult to obtain organisms using both culture-based and single-cell approaches.

A catalogue of all HMP reference genomes along with custom filtering, viewing, graphing and download options can be found at the DACC Project Catalogue (http://www.hmpdacc-resources.org/hmp_catalog/main.cgi). In addition, comparative analyses of reference genomes are provided by the data warehouse and analytical systems, Integrated Microbial Genomes/HMP (http://www.hmpdacc-resources.org/cgi-bin/imgm_hmp/main.cgi). Cultures of all HMP reference strains are required to be made publicly available through the Biodefense and Emerging Infections Research Resources Repository (BEI). Information on strain acquisition can be found at the DACC (http://hmpdacc.org/reference_genomes/reference_genomes.php) and BEI (http://www.beiresources.org/tabid/1901/stabid/1901/CollectionLinkID/4/Default.aspx).

Conclusion

An overarching goal of this multi-year, multi-centre project is the generation of a community resource to advance research efforts related to the microbiome. The result is a collection of 11,174 primary biological specimens representing the human microbiome, as well as corresponding blood samples from the human donors, which are being reserved for sequencing at a future date and from which cell lines will be developed. A variety of new protocols were developed to enable a project of this scope; these include methods for donor recruitment, laboratory and sequence processing, and analysis of

16S and WGS sequence and profiles. These resources serve as models to guide the design of similar projects. Studies with a primary focus on disease can use this reference for comparative purposes, including detecting shifts in microbial taxonomic and functional profiles, or identification of new species not present in healthy cohorts that appear under disease conditions. The catalogue described in this study is, to our knowledge, the largest and most comprehensive reference set of human microbiome data associated with healthy adult individuals. Collectively the data represent a treasure trove that can be mined to identify new organisms, gene functions, and metabolic and regulatory networks, as well as correlations between microbial community structure and health and disease[4]. Among other future benefits, this resource may promote the development of novel prophylactic strategies such as the application of prebiotics and probiotics to foster human health.

METHODS SUMMARY

As part of a multi-institutional collaboration, the HMP human subjects study was reviewed by the Institutional Review Boards (IRBs) at each sampling site: the BCM (IRB protocols H-22895 (IRB no. 00001021) and H-22035 (IRB no. 00002649)); Washington University School of Medicine (IRB protocol HMP-07-001 (IRB no. 201105198)); and St Louis University (IRB no. 15778). The study was also reviewed by the J. Craig Venter Institute under IRB protocol 2008-084 (IRB no. 00003721), and at the Broad Institute of MIT and Harvard the study was determined to be exempt from IRB review. All study participants gave their written informed consent before sampling and the study was conducted using the Human Microbiome Project Core Sampling Protocol A. Each IRB has a federal-wide assurance and follows the regulations established in 45 CFR Part 46. The study was conducted in accordance with the ethical principles expressed in the Declaration of Helsinki and the requirements of applicable federal regulations.

All further details are in Supplementary Information.

Received 2 November 2011; accepted 10 May 2012.

1. Gilbert, J. A. & Dupont, C. L. Microbial metagenomics: beyond the genome. *Annu. Rev. Mar. Sci.* **3,** 347–371 (2011).
2. NIH HMP Working Group et al. The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–2323 (2009).
3. Human Microbiome Jumpstart Reference Strains Consortium. A catalog of reference genomes from the human microbiome. *Science* **328,** 994–999 (2010).
4. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* http://dx.doi.org/10.1038/nature11234 (this issue).
5. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
6. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* http:dx.plos.org/10.1371/journal.pone.0039315 (14 June 2012).
7. Kunin, V., Engelbrektson, A., Ochman. H. & Hugenholtz, P. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12,** 118–123 (2010).
8. Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L. & Welch, D. M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8,** R143 (2007).
9. Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6,** e27310 (2011).
10. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).
11. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
12. Martin, J. S. *et al.* Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.0036427 (14 June 2012).
13. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* http://dx.doi.org/10.1371/journal.pcbi.1002358 (14 June 2012).
14. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nature Genet.* **25,** 25–29 (2000).
15. Goll, J. *et al.* A case study for large-scale human microbiome analysis using JCVI's Metagenomics Reports (METAREP). *PLoS ONE* http://dx.doi.org/10.1371/journal.pone.002904 (14 June 2012).
16. Muller, J. *et al.* eggNOG v2.0: extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res.* **38,** D190–D195 (2010).
17. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The consortium would like to thank our external scientific advisory board: R. Blumberg, J. Davies, R. Holt, P. Ossorio, F. Ouellette, G. Schoolnik and A. Williamson. We would also like to thank our collaborators throughout the International Human Microbiome Consortium, particularly the investigators of the MetaHIT project, for advancing human microbiome research. Data repository management was provided by the NCBI and the Intramural Research Program of the NIH National Library of Medicine. We especially appreciate the generous participation of the individuals from the St Louis, Missouri, and Houston, Texas areas who made this study possible. This research was supported in part by NIH grants U54HG004969 to B.W.B.; U54HG003273 to R.A.G.; U54HG004973 to R.A.G., S.K.H. and J.F.P.; U54HG003067 to E. S. Lander.; U54AI084844 to K.E.N.; N01AI30071 to R. L. Strausberg; U54HG004968 to G.M.W.; U01HG004866 to O.W.; U54HG003079 to R.K.W.; R01HG005969 to C.H.; R01HG004872 to R.K.; R01HG004885 to M.P.; R01HG005975 to P.D.S.; R01HG004908 to Y.Y.; R01HG004900 to M. K. Cho and P. Sankar; R01HG005171 to D.E.H.; R01HG004853 to A.L.M.; R01HG004856 to R.R.; R01HG004877 to R.R.S. and R.M.F.; R01HG005172 to P. Spicer; R01HG004857 to M.P.; R01HG004906 to T.M.S.; R21HG005811 to E.A.-V.; G.A.B. was supported by UH2AI083263 and UH3AI083263 (G.A.B., C. N. Cornelissen, L. K. Eaves and J. F. Strauss); M.J.B. was supported by UH2AR057506, S.M.H. was supported by UH3DK083993 (V. B. Young, E. B. Chang, F. Meyer, T.M.S., M. L. Sogin, J. M. Tiedje); K.P.R. was supported by UH2DK083990 (J.V.); J.A.S. and H.H.K. were supported by UH2AR057504 and UH3AR057504 (J.A.S.); DP2OD001500 to K.M.A.; N01HG62088 to the Coriell Institute for Medical Research; U01DE016937 to F.E.D.; S.K.-H. was supported by RC1DE020298and R01DE021574 (S.K.-H. and H. Li); J.I. was supported by R21CA139193 (J.I. and D. S. Michaud); K.P.L. was supported by P30DE020751 (D. J. Smith); Army Research Office grant W911NF-11-1-0473 to C.H.; National Science Foundation grants NSF DBI-1053486 to C.H. and NSF IIS-0812111 to M.P.; The Office of Science of the US Department of Energy under contract no. DE-AC02-05CH11231 for P.S.C.; LANL Laboratory-Directed Research and Development grant 20100034DR and the US Defense Threat Reduction Agency grants B104153I and B084531I to P.S.C.; Research Foundation - Flanders (FWO) grant to K.F. and J. Raes; R.K. is a Howard Hughes Medical Institute (HHMI) Early Career Scientist; Gordon & Betty Moore Foundation funding and institutional funding from the J. David Gladstone Institutes to K.S.P.; A.M.S. was supported by fellowships provided by the Rackham Graduate School and the NIH Molecular Mechanisms in Microbial Pathogenesis Training Grant T32AI007528; a Crohn's and Colitis Foundation of Canada Grant in Aid of Research to E.A.-V.; 2010 IBM Faculty Award to K.C.W. Analysis of the HMP data was performed using National Energy Research Scientific Computing resources; the BluBioU Computational Resource at Rice University.

Author Contributions Principal investigators: B.W.B., R.A.G., S.K.H., B.A.M., K.E.N., J.F.P., G.M.W., O.W., R.K.W. Manuscript preparation: B.A.M., K.E.N., M.P., H.H.C., M.G.G., D.G., C.H., J.F.P. Funding agency management: C.C.B., T.B., V.R.B., J.L.C., S.C., C.D., V.D.F., C.G., M.Y.G., R.D.L., J.M., P.M., J.P., L.M.P., J.A.S., L.W., C.W., K.A.W. Project leadership: S.A., J.H.B., B.W.B., A.T.C., H.H.C., A.M.E., M.G.F., R.S.F., D.G., M.G.G., K.H., S.K.H., C.H., E.A.L., R.M., V.M., J.C.M., B.A.M., M.M., D.M.M., K.E.N., J.F.P., E.J.S., J.V., G.M.W., O.W., A.M.W., K.C.W., J.R.W., S.K.Y., Q.Z. Analysis preparation for manuscript: M.B., B.L.C., D.G., M.G.G., M.E.H., C.H., K.L., B.A.M., X.Q., J.R.W., M.T. Data release: L.A., T.B., I.A.C., K.C., H.H.C., N.J.D., D.J.D., A.M.E., V.M.F., L.F., J.M.G., S.G., S.K.H., M.E.H., C.J., V.J., C.K., A.A.M., V.M.M., T.M., M.M., D.M.M., J.O., K.P., J.F.P., C.P., X.Q., R.K.S., N.S., I.S., E.J.S., D.V.W., O.W., K.W., K.C.W., C.Y., B.P.Y., Q.Z. Methods and research development: S.A., H.M.A., M.B., D.M.C., A.M.E., R.L.E., M.F., S.F., M.G.F., D.C.F., D.G., G.G., B.J.H., S.K.H., M.E.H., W.A.K., N.L., K.L., V.M., E.R.M., B.A.M., M.M., D.M.M., C.N., J.F.P., M.E.P., X.Q., M.C.R., C.R., E.J.S., S.M.S., D.G.T., D.V.W., G.M.W., Y.W., K.M.W., S.Y., B.P.Y., S.K.Y., Q.Z. DNA sequence production: S.A., E.A., T.A., T.B., C.J.B., D.A.B., K.D.D., S.P.D., A.M.E., R.L.E., C.N.F., S.F., C.C.F., L.L.F., R.S.F., B.H., S.K.H., M.E.H., V.J., C.L.K., S.L.L., N.L., L.L., D.M.M., I.N., C.N., M.O., J.F.P., X.Q., J.G.R., Y.R., M.C.R., D.V.W., Y.W., B.P.Y., Y.Z. Clinical sample collection: K.M.A., M.A.C., W.M.D., L.L.F., N.G., H.A.H., E.L.H., J.A.K., W.A.K., T.M., A.L.M., P.M., S.M.P., J.F.P., G.A.S., J.V., M.A.W., G.M.W. Body site experts: K.M.A., E.A.V., G.A., L.B., M.J.B., C.C.D., F.E.D., L.F., J.I., J.A.K., S.K.H., H.H.K., K.P.L., P.J.M., J. Ravel., T.M.S., J.A.S., J.D.S., J.V. Ethical, legal and social implications: R.M.F., D.E.H., W.A.K., N.B.K., C.M.L., A.L.M., R.R., P. Sankar, P. Spicer, R.R.S., L.Z. Strain management: E.A.V., J.H.B., I.A.C., K.C., S.W.C., H.H.C., T.Z.D., A.S.D., A.M.E., M.G.F., M.G.G., S.K.H., V.J., N.C.K., S.L.L., L.L., K.L., E.A.L., V.M.M., B.A.M., D.M.M., K.E.N., I.N., I.P., L.S., E.J.S., C.M.T., M.T., D.V.W., G.M.W., A.M.W., Y.W., K.M.W., B.P.Y., L.Z., Y.Z. 16S data analysis: K.M.A., E.J.A., G.L.A., C.A.A., M.B., B.W.B., J.P.B., G.A.B., S.R.C., S.C., J.C., T.Z.D., F.E.D., E.D., A.M.E., R.C.E., M.F., A.A.F., J.F., K.F., H.G., D.G., B.J.H., T.A.H., S.M.H., C.H., J.I., J.K.J., S.T.K., S.K.H., R.K., H.H.K., O.K., P.S.L., R.E.L., K.L., C.A.L., D.M., B.A.M., K.A.M., M.M., M.P., J.F.P., M.P., K.S.P., X.Q., J. Raes, K.P.R., M.C.R., B.R., J.F.S., P.D.S., T.M.S., N.S., J.A.S., W.D.S., T.J.S., C.S.S., E.J.S., R.M.T., J.V., T.A.V., Z.W., D.V.W., G.M.W., J.R.W., K.M.W., Y.Y., S.Y., Y.Z. Shotgun data processing and alignments: C.J.B., J.C.C., E.D., D.G., A.G., M.E.H., H.J., D.K., K.C.K., C.L.K., Y.L., J.C.M., B.A.M., M.M., D.M.M., J.O., J.F.P., X.Q., J.G.R., R.K.S., N.U.S., I.S., E.J.S., G.G.S., S.M.S., J.W., Z.W., G.M.W., O.W., K.C.W., T.W., S.K.Y., L.Z. Assembly: H.M.A., C.J.B., P.S.C., L.C., Y.D., S.P.D., M.G.F., M.E.H., H.J., S.K., B.L., Y.L., C.L., J.C.M., J.M.M., J.R.M., P.J.M., M.M., J.F.P., M.P., M.E.P., X.Q., M.R., R.K.S., M.S., D.D.S., G.G.S., S.M.S., C.M.T., T.J.T., W.W., G.M.W., K.C.W., L.Y., Y.Y., S.K.Y., L.Z. Annotation: O.O.A., V.B., C.J.B., I.A.C., A.T.C., K.C., H.H.C., A.S.D., M.G.G., J.M.G., J.G., A.G., S.G., B.J.H., K.H., S.K.H., C.H., H.J., N.C.K., R.M., V.M.M., K.M., T.M., M.M., J.O., K.P., M.P., X.Q., N.S., E.J.S., G.G.S., S.M.S., M.T., G.M.W., K.C.W., J.R.W., C.Y., S.K.Y., Q.Z., L.Z. WGS Metabolic Reconstruction: S.A., B.L.C., J.G., C.H., J.I., B.A.M., M.M., B.R., A.M.S., N.S., M.T., G.M.W., S.Y., Q.Z., J.D.Z.

Author Information Accession numbers for all primary sequencing data are given in Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share-Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial

interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.A.M. (bmethe@jcvi.org).

The Human Microbiome Project Consortium

Barbara A. Methé[1], Karen E. Nelson[1], Mihai Pop[2], Heather H. Creasy[3], Michelle G. Giglio[3], Curtis Huttenhower[4,5], Dirk Gevers[5], Joseph F. Petrosino[6,15,79], Sahar Abubucker[7], Jonathan H. Badger[77], Asif T. Chinwalla[7], Ashlee M. Earl[5], Michael G. FitzGerald[5], Robert S. Fulton[7], Kymberlie Hallsworth-Pepin[7], Elizabeth A. Lobos[7], Ramana Madupu[1], Vincent Magrini[7], John C. Martin[7], Makedonka Mitreva[7], Donna M. Muzny[6], Erica J. Sodergren[7], James Versalovic[8,9], Aye M. Wollam[7], Kim C. Worley[6], Jennifer R. Wortman[5], Sarah K. Young[5], Qiandong Zeng[5], Kjersti M. Aagaard[10], Olukemi O. Abolude[3], Emma Allen-Vercoe[11], Eric J. Alm[5,12], Lucia Alvarado[5], Gary L. Andersen[13], Scott Anderson[5], Elizabeth Appelbaum[7], Harindra M. Arachchi[5], Gary Armitage[14], Cesar A. Arze[3], Tulin Ayvaz[15], Carl C. Baker[16], Lisa Begg[17], Tsegahiwot Belachew[18], Veena Bhonagiri[7], Monika Bihan[1], Martin J. Blaser[19], Toby Bloom[5], Vivien R. Bonazzi[20], Paul Brooks[21,22], Gregory A. Buck[22,23], Christian J. Buhay[6], Dana A. Busam[1], Joseph L. Campbell[18,20], Shane R. Canon[24], Brandi L. Cantarel[3], Patrick S. Chain[25,26], I-Min A. Chen[27], Lei Chen[7], Shaila Chhibba[20], Ken Chu[27], Dawn M. Ciulla[5], Jose C. Clemente[28], Sandra W. Clifton[7], Sean Conlan[80], Jonathan Crabtree[3], Mary A. Cutting[29], Noam J. Davidovics[3], Catherine C. Davis[30], Todd Z. DeSantis[31], Carolyn Deal[18], Kimberley D. Delehaunty[7], Floyd E. Dewhirst[32,33], Elena Deych[7], Yan Ding[6], David J. Dooling[7], Shannon P. Dugan[6], W. Michael Dunne Jr[34,35], A. Scott Durkin[1], Robert C. Edgar[36], Rachel L. Erlich[5], Candace N. Farmer[7], Ruth M. Farrell[37], Karoline Faust[38,39], Michael Feldgarden[5], Victor M. Felix[3], Sheila Fisher[5], Anthony A. Fodor[40], Larry Forney[41], Leslie Foster[1], Valentina Di Francesco[18], Jonathan Friedman[42], Dennis C. Friedrich[5], Catrina C. Fronick[7], Lucinda L. Fulton[7], Hongyu Gao[8], Nathalia Garcia[43], Georgia Giannoukos[5], Christina Giblin[18], Maria Y. Giovanni[18], Jonathan M. Goldberg[5], Johannes Goll[1], Antonio Gonzalez[44], Allison Griggs[5], Sharvari Gujja[5], Brian J. Haas[5], Holli A. Hamilton[29], Emily L. Harris[29], Theresa A. Hepburn[5], Brandi Herter[7], Diane E. Hoffmann[45], Michael E. Holder[6], Clinton Howarth[5], Katherine H. Huang[5], Susan M. Huse[46], Jacques Izard[32,47], Janet K. Jansson[48], Huaiyang Jiang[6], Catherine Jordan[3], Vandita Joshi[6], James A. Katancik[49], Wendy A. Keitel[15], Scott T. Kelley[50], Cristyn Kells[5], Susan Kinder-Haake[51]†, Nicholas B. King[52], Rob Knight[28,53], Dan Knights[44], Heidi H. Kong[54], Omry Koren[55], Sergey Koren[2], Karthik C. Kota[7], Christie L. Kovar[6], Nikos C. Kyrpides[26], Patricio S. La Rosa[56], Sandra L. Lee[6], Katherine P. Lemon[32,57], Niall Lennon[6], Cecil M. Lewis[58], Lora Lewis[6], Ruth E. Ley[55], Kelvin Li[1], Konstantinos Liolios[26], Bo Liu[2], Yue Liu[6], Chien-Chi Lo[25], Catherine A. Lozupone[28], R. Dwayne Lunsford[29], Tessa Madden[59], Anup A. Mahurkar[3], Peter J. Mannon[60], Elaine R. Mardis[7], Victor M. Markowitz[26,27], Konstantinos Mavrommatis[26], Jamison M. McCorrison[1], Daniel McDonald[28], Jean McEwen[20], Amy L. McGuire[61], Pamela McInnes[29], Teena Mehta[5], Kathie A. Mihindukulasuriya[7], Jason R. Miller[1], Patrick J. Minx[7], Irene Newsham[6], Chad Nusbaum[5], Michelle O'Laughlin[7], Joshua Orvis[3], Ioanna Pagani[26], Krishna Palaniappan[27], Shital M. Patel[62], Matthew Pearson[5], Jane Peterson[20], Mircea Podar[63], Craig Pohl[7], Katherine S. Pollard[64,65,66], Margaret E. Priest[5], Lita M. Proctor[20], Xiang Qin[6], Jeroen Raes[38,39], Jacques Ravel[3,67], Jeffrey G. Reid[6], Mina Rho[68], Rosamond Rhodes[69], Kevin P. Riehle[70], Maria C. Rivera[22,23], Beltran Rodriguez-Mueller[50], Yu-Hui Rogers[1], Matthew C. Ross[15], Carsten Russ[5], Ravi K. Sanka[1], Pamela Sankar[71], J. Fah Sathirapongsasuti[4], Jeffery A. Schloss[20], Patrick D. Schloss[72], Thomas M. Schmidt[73], Matthew Scholz[25], Lynn Schriml[3], Alyxandria M. Schubert[72], Nicola Segata[4], Julia A. Segre[80], William D. Shannon[56], Richard R. Sharp[37], Thomas J. Sharpton[64], Narmada Shenoy[5], Nihar U. Sheth[22], Gina A. Simone[74], Indresh Singh[1], Chris S. Smillie[42], Jack D. Sobel[75], Daniel D. Sommer[2], Paul Spicer[58], Granger G. Sutton[1], Sean M. Sykes[5], Diana G. Tabbaa[5], Mathangi Thiagarajan[1], Chad M. Tomlinson[7], Manolito Torralba[1], Todd J. Treangen[76], Rebecca M. Truty[64], Tatiana A. Vishnivetskaya[63], Jason Walker[7], Lu Wang[20], Zhengyuan Wang[5], Doyle V. Ward[5], Wesley Warren[7], Mark A. Watson[34], Christopher Wellington[20], Kris A. Wetterstrand[20], James R. White[3], Katarzyna Wilczek-Boney[6], Yuan Qing Wu[6], Kristine M. Wylie[7], Todd Wylie[7], Chandri Yandava[5], Liang Ye[7], Yuzhen Ye[68], Shibu Yooseph[77], Bonnie P. Youmans[15], Lan Zhang[6], Yanjiao Zhou[7], Yiming Zhu[6], Laurie Zoloth[78], Jeremy D. Zucker[5], Bruce W. Birren[5], Richard A. Gibbs[6], Sarah K. Highlander[6,15], George M. Weinstock[7], Richard K. Wilson[7] & Owen White[3]

[1]J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, Maryland 20850, USA. [2]University of Maryland, Center for Bioinformatics and Computational Biology and Department of Computer Science, Biomolecular Sciences Building Rm. 3120F, College Park, Maryland 20742, USA. [3]University of Maryland School of Medicine, Institute for Genome Sciences 801 W. Baltimore Street, Baltimore, Maryland 21201, USA. [4]Harvard School of Public Health, Department of Biostatistics, 655 Huntington Avenue, Boston, Massachusetts 02115, USA. [5]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [6]Baylor College of Medicine Human Genome Sequencing Center, One Baylor Plaza, Houston, Texas 77030, USA. [7]Washington University School of Medicine, The Genome Institute, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [8]Baylor College of Medicine, Department of Pathology & Immunology, One Baylor Plaza, Houston, Texas 77030, USA. [9]Texas Children's Hospital Department of Pathology, 6621 Fannin Street, Houston, Texas 77030, USA. [10]Baylor College of Medicine, Department of Obstetrics & Gynecology, Division of Maternal-Fetal Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [11]University of Guelph Department of Molecular and Cellular Biology, 50 Stone Road East, Guleph, Ontario N1G 2W1, Canada. [12]Massachusetts Institute of Technology, Department of Civil & Environmental Engineering, Parsons Laboratory, Room 48-317, 15 Vassar Street, Cambridge, Massachusetts 02139, USA. [13]Lawrence Berkeley National Laboratory, Center for Environmental Biotechnology, 1 Cyclotron Road, Berkeley, California 94720, USA. [14]University of California, San Francisco, School of Dentistry, 707 Parnassus Avenue, San Francisco, California 94143, USA. [15]Baylor College of Medicine, Molecular Virology and Microbiology, One Baylor Plaza, Houston, Texas 77030, USA. [16]National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), 6701 Democracy Boulevard, MSC 4872, Bethesda, Maryland 20892, USA. [17]National Institutes of Health, Office of Research on Women's Health (ORWH), 6707 Democracy Boulevard, MSC 5484, Bethesda, Maryland 20892, USA. [18]National Institutes of Health, National Institute for Allergy and Infectious Diseases (NIAID), 6610 Rockledge Drive, MSC 6603, Bethesda, Maryland 20892, USA. [19]New York University Langone Medical Center, Department of Medicine, 550 First Avenue, OBV A-606, New York, New York 10016, USA. [20]National Institutes of Health, National Human Genome Research Institute (NHGRI), 5635 Fishers Lane, MSC 9305, Bethesda, Maryland 20892, USA. [21]Virginia Commonwealth University, Department of Statistical Sciences and Operations Research, PO Box 843083, Richmond, Virginia 23284, USA. [22]Virginia Commonwealth University, Center for the Study of Biological Complexity, 1000 West Cary Street, Richmond, Virginia 23284, USA. [23]Virginia Commonwealth University, Department of Biology, 1000 West Cary Street, Richmond, Virginia 23284, USA. [24]Lawrence Berkeley National Laboratory, Technology Integration Group, National Energy Research Scientific Computing Center, 1 Cyclotron Road, Berkeley, California 94720, USA. [25]Los Alamos National Laboratory Genome Science Group, Bioscience Division, HRL, MS-888, LANL, Los Alamos, New Mexico 87545, USA. [26]Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. [27]Lawrence Berkeley National Laboratory, Biological Data Management and Technology Center, Computational Research Division, 1 Cyclotron Road, Berkeley, California 94720, USA. [28]University of Colorado, Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado, Boulder, Colorado 80309-0215, USA. [29]National Institutes of Health, National Institute of Dental and Craniofacial Research (NIDCR), 6701 Democracy Boulavard, MSC 4878, Bethesda, Maryland 20892, USA. [30]The Procter & Gamble Company, FemCare Product Safety and Regulatory Affairs, 6110 Center Hill Avenue, Cincinnati, Ohio 45224, USA. [31]Second Genome, Inc. Bioinformatics Department, 1150 Bayhill Drive, Suite 215, San Bruno, California 94066, USA. [32]Forsyth Institute, Department of Molecular Genetics, 245 First Street, Cambridge, Massachusetts 02142, USA. [33]Harvard School of Dental Medicine, Department of Oral Medicine, Infection and Immunity, 188 Longwood Avenue, Boston, Massachusetts 02115, USA. [34]Washington University School of Medicine, Department of Pathology & Immunology, 660 South Euclid Avenue, Box 8118, St Louis, Missouri 63110, USA. [35]bioMerieux, Inc., 100 Rodolphe Street, Durham, North Carolina 27712, USA. [36]drive5.com, Tiburon, California 94920, USA. [37]Cleveland Clinic, Center for Bioethics, Humanities and Spiritual Care, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA. [38]VIB, Belgium, Department of Structural Biology, Pleinlaan 2, 1050 Brussels, Belgium. [39]Vrije Universiteit Brussels, Department of Applied Biological Sciences (DBIT), Pleinlaan 2, 1050 Brussels, Belgium. [40]University of North Carolina Charlotte, Department of Bioinformatics and Genomics, 9201 University City Blvd, Charlotte, North Carolina 28223-0001, USA. [41]University of Idaho, Department of Biological Sciences, Life Sciences South Room 441A, PO Box 443051, Moscow, Idaho 83844, USA. [42]Massachusetts Institute of Technology, Computational and Systems Biology, Parsons Laboratory, Room 48-317, 15 Vassar Street, Cambridge, Massachusetts 02139, USA. [43]Saint Louis University, Center for Advanced Dental Education, 3320 Rutger Street, St Louis, Missouri 63104, USA. [44]University of Colorado, Department of Computer Science, University of Colorado, Boulder, Colorado 80309, USA. [45]University of Maryland Francis King Carey School of Law, 500 W. Baltimore Street, Baltimore, Maryland 21201, USA. [46]Marine Biological Laboratory, Josephine Bay Paul Center, 7 MBL Street, Woods Hole, Massachusetts 02543-1015, USA. [47]Harvard School of Dental Medicine, Department of Oral Medicine, Infection and Immunity, 188 Longwood Avenue, Boston, Massachusetts 02115, USA. [48]Lawrence Berkeley National Laboratory, Ecology Department, Earth Sciences Division, 1 Cyclotron Road, Berkeley, California 94720, USA. [49]University of Texas Health Science Center School of Dentistry, Department of Periodontics, 6516 MD Anderson Blvd, Houston, Texas 77030, USA. [50]San Diego State University, Department of Biology, 5500 Campanile Drive, San Diego, California 92182, USA. [51]UCLA School of Dentistry, Division of Associated Clinical Specialties and Dental Research Institute, 10833 Le Conte Avenue, Los Angeles, California 90095-1668, USA. [52]McGill University, Faculty of Medicine, Peel 3647 Montreal, Quebec H3A 1X1, Canada. [53]Howard Hughes Medical Institute, Campus Box 215, Boulder, Colorado 80309-0215, USA. [54]National Institutes of Health, National Cancer Institute (NCI), Dermatology Branch, CCR, MSC 1908, 10 Center Drive, Bethesda, Maryland 20892, USA. [55]Cornell University, Department of Microbiology, 467 Biotechnology Building, Ithaca, New York 14853, USA. [56]Washington University School of Medicine, Department of Medicine, Division of General Medical Science, 660 South Euclid Avenue, Box 8005, St Louis, Missouri 63110, USA. [57]Children's Hospital Boston, Harvard Medical School, Division of Infectious Diseases, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [58]University of Oklahoma, Department of Anthropology, 455 West Lindsey, Dale Hall Tower 521, Norman, Oklahoma 73019, USA. [59]Washington University School of Medicine, Department of Obstetrics and Gynecology, 4533 Clayton Avenue, Box 8219, St Louis, Missouri 63110, USA. [60]University of Alabama at Birmingham, Division of Gastroenterology and Hepatology, 1530 3rd Avenue South, Birmingham, Alabama 35294-1150, USA. [61]Baylor College of Medicine, Center for Medical Ethics and Health Policy, One Baylor Plaza, Houston, Texas 77030, USA. [62]Baylor College of Medicine, Medicine-Infectious Disease, One Baylor Plaza, Houston, Texas 77030, USA. [63]Oak Ridge National Laboratory, Biosciences Division, PO Box 2008 MS 6038 Oak Ridge, Tennessee 37831-6038, USA. [64]University of California, San Francisco, Gladstone Institutes, 1650 Owens Street, San Francisco, California 94158, USA. [65]University of California, San Francisco, Institute for Human Genetics, 1650 Owens Street, San Francisco, California 94158, USA. [66]University of California, San Francisco, Division of Biostatistics, 1650 Owens Street, San Francisco, California 94158, USA. [67]University of Maryland School of Medicine, Department of Microbiology and Immunology, BioPark II - Room 611, 801 W. Baltimore Street, Baltimore, Maryland 21201, USA. [68]Indiana University, School of Informatics and Computing, 150 S. Woodlawn Avenue, Bloomington, Indiana 47405, USA. [69]Mount Sinai School of Medicine, Annenberg Building Floor 5th, Room 5-208, 1468 Madison Avenue, New York, New York 10029, USA. [70]Baylor College of Medicine Molecular & Human Genetics, One Baylor Plaza, Houston,

Texas 77030, USA. [71]University of Pennsylvania, Center for Bioethics and Department of Medical Ethics, 3401 Market Street, Suite 320, Philadelphia, Pennsylvania 19104, USA. [72]University of Michigan, Department of Microbiology & Immunology, 5713 Medical Science Bldg. II, 1150 West Medical Center Dr., Ann Arbor, Michigan 48109-5620, USA. [73]Michigan State University, Department of Microbiology and Molecular Genetics, 6180 Biomedical Physical Sciences, Michigan State University, East Lansing, Michigan 48824, USA. [74]The EMMES Corporation, 401 N. Washington St., Suite 700, Rockville, Maryland 20850, USA. [75]Wayne State University School of Medicine, Detroit, Michigan, Harper University Hospital, 3990 John R Street, Detroit, Michigan 48201, USA. [76]Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine, Bloomberg School of Public Health, E3138, 615 N Wolfe St, Baltimore, Maryland 21205, USA. [77]J. Craig Venter Institute, 10355 Science Center Drive, San Diego, California 92121, USA. [78]Northwestern University, Feinberg School of Medicine, 420 East Superior Street Chicago, Illinois 60611, USA. [79]Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [80]National Institutes of Health, National Human Genome Research Institute (NHGRI), Genetics and Molecular Biology Branch, MSC 4442, Bethesda, Maryland 20892, USA.
†Deceased.

ARTICLE

doi:10.1038/nature11053

Human gut microbiome viewed across age and geography

Tanya Yatsunenko[1], Federico E. Rey[1], Mark J. Manary[2,3], Indi Trehan[2,4], Maria Gloria Dominguez-Bello[5], Monica Contreras[6], Magda Magris[7], Glida Hidalgo[7], Robert N. Baldassano[8], Andrey P. Anokhin[9], Andrew C. Heath[9], Barbara Warner[2], Jens Reeder[10], Justin Kuczynski[10], J. Gregory Caporaso[11], Catherine A. Lozupone[10], Christian Lauber[10], Jose Carlos Clemente[10], Dan Knights[10], Rob Knight[10,12] & Jeffrey I. Gordon[1]

Gut microbial communities represent one source of human genetic and metabolic diversity. To examine how gut microbiomes differ among human populations, here we characterize bacterial species in fecal samples from 531 individuals, plus the gene content of 110 of them. The cohort encompassed healthy children and adults from the Amazonas of Venezuela, rural Malawi and US metropolitan areas and included mono- and dizygotic twins. Shared features of the functional maturation of the gut microbiome were identified during the first three years of life in all three populations, including age-associated changes in the genes involved in vitamin biosynthesis and metabolism. Pronounced differences in bacterial assemblages and functional gene repertoires were noted between US residents and those in the other two countries. These distinctive features are evident in early infancy as well as adulthood. Our findings underscore the need to consider the microbiome when evaluating human development, nutritional needs, physiological variations and the impact of westernization.

Genetic variation between human populations is typically viewed as differences in the allele frequencies of shared *Homo sapiens* genes. Another source of genetic and metabolic diversity resides in differences in the representation of the millions of genes and myriad gene functions within our gut microbial communities[1–3]. Sampling a broad population of healthy humans representing different ages and cultural traditions offers an opportunity to discover how our gut microbiomes evolve within a lifespan, vary between populations, and respond to our changing lifestyles[1,4–9]. Therefore, we conducted a demonstration project to address the question of whether there are discernible patterns of functional maturation of the gut communities of healthy infants and children living in geographically and culturally distinct settings.

Fecal samples were obtained from individuals in families of Guahibo Amerindians residing in two villages (Platanillal and Coromoto), separated by 10 miles, and located near Puerto Ayacucho in the Amazonas State of Venezuela (see Supplementary Table 1a, b for information about their diets). Fecal samples were also procured from members of families living in four rural communities of Malawi located within 10–70 miles of one another (Chamba, Makwhira, Mayaka and Mbiza). Lifestyles in these villages are very similar, and diets are relatively monotonous, dominated by maize (Supplementary Table 1c). In addition, we sampled families distributed across the United States, including the greater metropolitan areas of St Louis, Philadelphia and Boulder. The sampled populations included parents and siblings, and, in the United States and Malawi, monozygotic and dizygotic twin pairs. A total of 531 individuals (151 families) were studied: 115 individuals (34 families) from Malawi; 100 individuals (19 families) from Venezuela; and 316 individuals (98 families) from the United States (see Supplementary Table 2 for subject characteristics; note that all except 35 adults and one child from the United States were explicitly recruited for this study).

DNA was extracted from a single fecal sample donated by each person. Variable region 4 (V4) of bacterial 16S ribosomal RNA genes present in each fecal community was amplified by PCR, and the resulting amplicons were sequenced on an Illumina HiSeq 2000 instrument ($n = 1{,}803{,}250 \pm 562{,}877$ (mean $\pm$ s.d.) reads per fecal sample; 1,093,740,274 total reads; Supplementary Table 2a) to define the phylogenetic types (phylotypes) present. Species-level bacterial phylotypes were defined as organisms sharing $\geq$97% nucleotide sequence identity in the V4 regions of their 16S rRNA genes[10]. In addition, we characterized functions encoded in community DNA by performing multiplex shotgun 454 pyrosequencing of fecal DNA from a subset of 110 fecal samples, encompassing 43 families with members matched as closely as possible for age ($155{,}890 \pm 87{,}083$ reads per sample; total size of data set, 5.9 Gb; Supplementary Table 2b). The resulting shotgun reads were annotated with Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology group (KO) assignments and with Enzyme Commision (EC) numbers (KEGG version 58).

Taxonomic changes as a function of age and population

Many reports have examined the bacterial species content of the gastrointestinal tracts of infants and children within one population using culture-based methods. Far fewer studies have attempted to compare the gut communities of humans living in markedly different socio-economic, geographic and cultural settings[11,12]. Culture-independent techniques have been used to define the gut microbiota at various points in postnatal development[6,13], but have been limited

[1]Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63108, USA. [2]Department of Pediatrics, Washington University School of Medicine, St Louis, Missouri 63110, USA. [3]Department of Community Health, University of Malawi College of Medicine, Blantyre, Malawi. [4]Department of Paediatrics and Child Health, University of Malawi College of Medicine, Blantyre, Malawi. [5]Department of Biology, University of Puerto Rico - Rio Piedras, Puerto Rico 00931-3360. [6]Venezuelan Institute of Scientific Research (IVIC), Carretera Panamericana, Km 11, Altos de Pipe, Venezuela. [7]Amazonic Center for Research and Control of Tropical Diseases (CAICET), Puerto Ayacucho 7101, Amazonas, Venezuela. [8]Division of Gastroenterology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA. [9]Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri 63110, USA. [10]Department of Chemistry and Biochemistry, University of Colorado, Boulder 80309, USA. [11]Department of Computer Science, Northern Arizona University, Flagstaff, Arizona 86001, USA. [12]Howard Hughes Medical Institute, University of Colorado, Boulder 80309, USA.

by the analytic methods used, by the low number of subjects examined, or by the scope of the populations surveyed. These studies have nonetheless provided important insights. Using 16S rRNA gene-based microarrays[14], a recent study found considerable intra- and interpersonal variation in fecal bacterial community structures during the first year of life in 12 unrelated children and 1 twin pair. Interpersonal variation was less within the twin pair, and intrapersonal variation decreased as a function of age. A quantitative PCR study of five bacterial taxa in the fecal microbiota of 1,032 Dutch infants at 1 month of age[15] documented differences based on birth mode (Caesarian versus vaginal; also see ref. 8).

We collected bacterial V4 16S rRNA data from 326 individuals aged 0–17 years (83 Malawian, 65 Amerindian and 178 US residents), plus 202 adults aged 18–70 years (31 Malawians, 35 Amerindians and 136 US residents). The 16S rRNA data sets were first analysed using UniFrac, an algorithm that measures similarity among microbial communities based on the degree to which their component taxa share branch length on a bacterial tree of life[16]. There were several notable findings. First, the phylogenetic composition of the bacterial communities evolved towards an adult-like configuration within the three-year period after birth in all three populations (Fig. 1a and Supplementary Fig. 1). Second, interpersonal variation was significantly greater among children than among adults; this finding was robust to geography (Fig. 1b; see also ref. 4). Third, there were significant differences in the phylogenetic composition of fecal microbiota between individuals living in the different countries, with especially pronounced separation occurring between the US and the Malawian and Amerindian gut communities; this was true for individuals aged 0–3 years, 3–17 years, and for adults (Fig. 1b and Supplementary Table 3). Unsupervised clustering using principal coordinates analysis (PCoA) of UniFrac distance matrices indicated that age and geography/cultural traditions primarily explain the variation in our data set, in which US microbiota clustered separately from non-US microbiota along principal coordinate 1 (Fig. 1c and Supplementary Fig. 2). However, within the non-US populations, separation between Malawians and Amerindians was also observed (along principal coordinate 3 in the case of adults; Supplementary Fig. 2f). We did not find any significant clustering by village for Malawians and Amerindians or by region within the United States. Fourth, bacterial diversity increased with age in all three populations (Fig. 2a, b). The fecal microbiota of US adults was the least diverse compared with the two other populations (Fig. 2c, $P < 0.005$, analysis of variance (ANOVA) with Bonferroni post-hoc test): these differences were evident in children older than 3 years of age ($P < 0.005$, ANOVA with Bonferroni post-hoc test), but not in younger subjects.

We next used the non-parametric Spearman rank correlation to determine which bacterial taxa change monotonically with increasing age within and between the three sampled populations. We only considered children who were breastfed and used data sets obtained from the V4 region of the 16S rRNA gene as well as data sets of shotgun pyrosequencing reads from the fecal microbiomes of the 110 sampled individuals (24 babies (0.6–5 months old), 60 children and adolescents (6 months to 17 years old) and 26 adults). Shotgun reads were mapped to 126 sequenced human gut-derived microbial species (Supplementary Table 4). The advantage of using these 126 gut microbes as a reference database is that spurious hits of shotgun microbiome reads to taxa that are not present in the gut are minimized. Nonetheless, when we repeated the entire analysis, blasting against 1,280 genomes in KEGG, the results were similar (Supplementary Fig. 3). Phylotypes belonging to *Bifidobacterium longum* exhibited a significant decline in proportional representation with increasing age in all three populations (Supplementary Fig. 3a). Most (75 ± 20%) shotgun and 16S rRNA V4 sequences in all babies mapped to members of the *Bifidobacterium* genus. Bifidobacteria continued to dominate fecal communities throughout the first year of life, although their

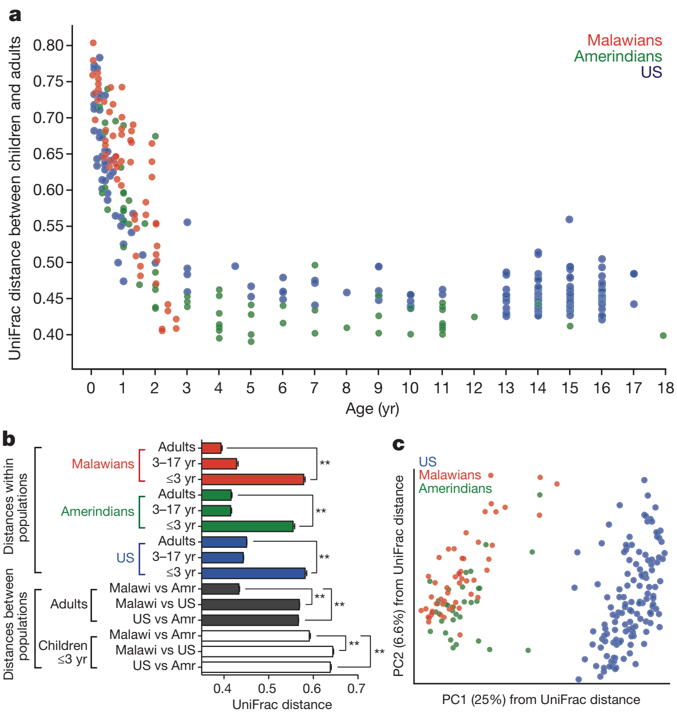


Figure 1 | Differences in the fecal microbial communities of Malawians, Amerindians and US children and adults. **a**, UniFrac distances between children and adults decrease with increasing age of children in each population. Each point shows the average distance between a child and all adults unrelated to that child but from the same country. Results are derived from bacterial V4 16S rRNA data sets. **b**, Large interpersonal variations are observed in the phylogenetic configurations of fecal microbial communities at early ages. Malawian and Amerindian (Amr) children and adults are more similar to one another than to US children and adults. UniFrac distances were defined from bacterial V4 16S rRNA data generated from the microbiota of 181 unrelated adults (≥18 years old) and 204 unrelated children ($n = 31$ Malawians 0.03–3 years old, 21 3–17 years old; 30 Amerindians 0.08–3 years old, 29 3–17 years old; 31 US residents 0.08–3 years old, 62 sampled at 3–17 years of age). $*P < 0.05$, $**P < 0.005$ (Student's *t*-test with 1,000 Monte Carlo simulations). See Supplementary Table 3 for a complete description of the statistical significance of all comparisons shown. **c**, PCoA of unweighted UniFrac distances for the fecal microbiota of adults. PC, principal coordinate.

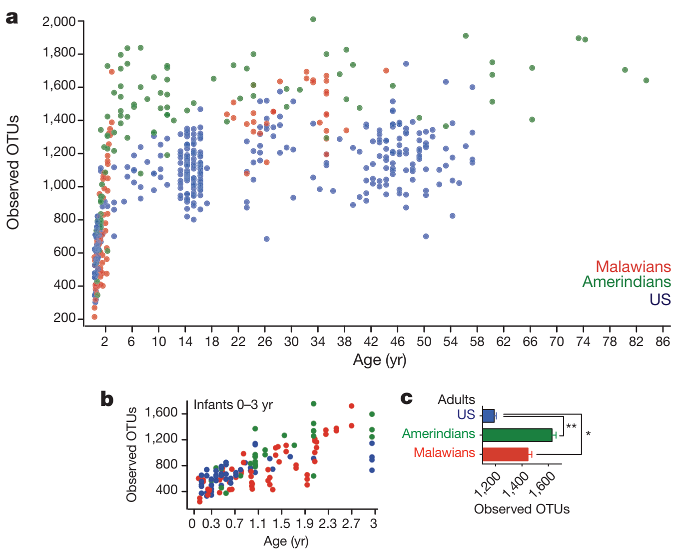


Figure 2 | Bacterial diversity increases with age in each population. **a–c**, The number of observed OTUs sharing ≥97% nucleotide sequence identity plotted against age for all subjects (**a**), during the first 3 years of life (**b**), and adults (**c**). Mean ± s.e.m. are shown in **c**. $*P < 0.05$, $**P < 0.005$ (ANOVA with Bonferroni post-hoc test).

proportional representation diminished during this period, in agreement with the results of several studies of small numbers of children[4,6,7] (Supplementary Fig. 3). Supplementary Table 5 lists the species-level bacterial taxa whose representation changes significantly with age in all three populations, as well as species that change in a population-specific manner as defined from analysis of the shotgun sequencing data that were available from 110 of the 531 individuals.

We also used Random Forests, a supervised machine-learning technique[17], and the V4 16S rRNA data sets obtained from 528 individuals to identify bacterial species-level operational taxonomic units (OTUs) that differentiate fecal community composition in children and adults within and between the three populations. The purpose of a classifier such as Random Forests is to learn a function that maps a set of input values or predictors (here, relative OTU abundances in a community) to a discrete output value (here, US versus non-US microbiota). Random Forests is a particularly powerful classifier that can exploit nonlinear relationships and complex dependencies between OTUs. The measure of the success of the method is its ability to classify unseen samples correctly, estimated by training it on a subset of samples, and using it to classify the remaining samples (cross-validation). The cross-validation error is compared with the baseline error that would be achieved by always guessing the most common category. As an added benefit, Random Forests assigns an importance score to each OTU by estimating the increase in error caused by removing that OTU from the set of predictors. In our analysis, we considered an OTU to be highly predictive if its importance score was at least 0.001; all error estimates and OTU importance scores were averaged over 100 rarefactions at the same sample size for each community (305,631 sequences) to control for sequencing effort.

Random Forests analysis confirmed the dominance of *Bifidobacterium* in the baby microbiota (Supplementary Table 6a). For adults, Random Forests revealed distinct community signatures for Western (US) and non-Western individuals (baseline error $= 0.289$, cross-validation error $= 0.011 \pm 0.000$). Of the 92 highly predictive species-level OTUs shown in Supplementary Table 6b, 73 were overrepresented in non-US adults, and 23 out of the 73 were assigned to the *Prevotella* genus. Malawians and Amerindians could also be distinguished from each other, although the difference was less extreme than the US versus non-US comparison (baseline error $= 0.407$, cross-validation error $= 0.018 \pm 0.009$, 56 highly predictive OTUs; Supplementary Table 6c). Only 28 OTUs distinguished US and non-US infants (Supplementary Table 6d). Intriguingly, three OTUs assigned to the *Prevotella* genus were overrepresented in the US infant microbiota, unlike the result observed in adults (Supplementary Table 6d). Twenty-three OTUs discriminated Malawian and Amerindian baby microbiomes, 20 of which were overrepresented in the latter: most belonged to the Enterococcaceae family (Supplementary Table 6d). Thus, a Western (US) lifestyle seems to affect the bacterial component of the gut microbiota systematically, although this influence is subtle compared with the high degree of variability observed in infants and children within each population (perhaps analogous to human genetic variability, in which variation among populations is small compared to variation within populations).

Confirming the importance of *Prevotella* as a discriminatory taxon, a recent study also showed that this genus was present in higher abundance in the fecal microbiota of children living in West Africa (Burkina Faso) compared with children living in Europe (Italy)[11]. Furthermore, a member of this genus is one of three bacterial species that, in European adults, distinguishes strongly among three clusters, or enterotypes, of gut microbiota configurations that are claimed to be reproducible across Western adult populations[18]. Therefore, we asked whether the fecal microbiota of infants and adults in each of our three geographically distinct populations fell into natural discrete clusters[19]. We did not find strong evidence for discrete clustering (see Methods), but rather for variation driven in adults by a trade-off between *Prevotella* and *Bacteroides* (Supplementary Fig. 4a). Including infants introduces a new, strongly supported gradient driven by *Bifidobacteria*, generally orthogonal to the *Bacteroides*/*Prevotella* trade-off (Supplementary Fig. 4b). Clustering of sub-populations of increasing minimum age indicates that adult cluster membership is generally consistent, but that children between 0.6 and 1 year of age may be clustered with adults or with younger children, depending on whether the younger children are included in the analysis (Supplementary Fig. 4c).

This clustering analysis suggests that some features of normal variation in the bacterial composition of the gut microbiota, such as the *Prevotella*/*Bacteroides* trade-off, are highly reproducible even in human population subsets of reduced variability. However, a complete description of variation in the human gut microbiota will require a substantially broader cross-cultural and cross-age sampling. Importantly, the observed age-related and geographic patterns were also detected with lower sequencing coverage (see Supplementary Results for a discussion of the influence of sequencing depth on the performance of Random Forests and PCoA analyses (Supplementary Figs 5 and 6), and our analysis of non-bacterial taxa that vary with age and population (Supplementary Fig. 7)).

Shared functional changes over time

Few studies have described changes in the gene content of the gut microbiome as a function of age: the largest study reported so far was carried out in 13 healthy Japanese individuals (5 children, the youngest 3 months old, plus 8 adults)[4]. Our shotgun sequencing data set from 110 individuals allowed us to characterize the representation of functional gene groups (KEGG KO annotations and EC numbers) in microbiomes representing broader age groups (youngest 3 weeks), and several distinct geographic locations and cultural traditions. We used Hellinger distance measurements to show that just as children are significantly more different from one another than are adults in terms of their fecal bacterial community phylogenetic structure, they are also more different in terms of their repertoires of microbiome-encoded functions, as defined by the proportional representation of EC and KO assignments. Moreover, as with UniFrac distances, Hellinger distances were greater between the US and the other two populations at all ages sampled (Supplementary Fig. 8). Of interest is the concordance of patterns of covariance between the two data types: Procrustes analysis disclosed that the goodness of fit was significant ($P < 0.001$ with 1,000 iterations) whether UniFrac (the most appropriate metric for 16S rRNA data) or Hellinger distances (for consistency with the method used on the KEGG EC and KEGG Orthology data) were used to reduce the OTU table (Supplementary Fig. 9a, b and data not shown). Annotation of shotgun reads from the microbiomes using the Clusters of Orthologous Groups (COG) database produced similar concordance with 16S rRNA data sets (Supplementary Fig. 9c).

When examining KEGG EC profiles across 110 fecal microbiomes, we obtained the remarkable result that there were no ECs identified as being unique to adults ($n = 26$) or babies (less than 6 months old, $n = 24$). Moreover, the total number of ECs found in adults was not significantly different than the total number of ECs scored in babies (sampling normalized to coverage in Supplementary Fig. 10a). This finding was robust to geography. The fraction of sequences with assignable KEGG EC annotations declined with increasing age in all three populations (Supplementary Fig. 10b). This may be due to the increased complexity of the adult microbiome, with fewer representative species characterized by genome sequencing, genetic manipulation or biochemically (also see Supplementary Results and Supplementary Figs 11 and 12 for a comparison of our data set to a published data set of fecal microbiomes sampled from 124 adults living in Denmark and Spain[2]).

We used ShotgunFunctionalizerR[20], a software tool designed for metagenomic analysis and based on a Poisson model, to identify 1,008 ECs whose proportional representation in fecal microbiomes differed significantly between all sampled breastfed babies and all adults irrespective of their geographic location; 530 were significantly higher

in adults ($P < 0.0001$, Supplementary Table 7). A prominent example of these shared age-related changes involves the metabolism of vitamins B12 (cobalamin) and folate. In contrast to folate, which is synthesized by microbes and plants, cobalamin is primarily produced by microbes[21]. The gut microbiomes of babies are enriched in genes involved in the *de novo* biosynthesis of folate, whereas those of adults have a significantly higher representation of genes that metabolize dietary folate and its reduced form tetrahydrofolate (THF; Supplementary Fig. 13 and Supplementary Table 7). Unlike *de novo* folate biosynthetic pathway components, which decrease with age, the proportional representation of genes encoding most enzymes involved in cobalamin biosynthesis increases with age (Supplementary Figs 14, 15 and Supplementary Table 7). The folate and cobalamin pathways are linked functionally by methionine synthase (EC2.1.1.13), which catalyses the formation of THF from 5-methyl-THF and L-homocysteine, requiring cobalamin as a cofactor; the representation of this enzyme also increases with age (Supplementary Fig. 13).

The low relative abundance of ECs involved in cobalamin biosynthesis in the fecal microbiomes of babies correlates with the lower representation of members of Bacteroidetes, Firmicutes and Archaea in their microbiota (see Supplementary Fig. 16 for Spearman correlation coefficients). Although the biosynthetic pathway for cobalamin is well represented in the genomes of these organisms (Supplementary Fig. 16), *Bifidobacterium*, *Streptococcus*, *Lactococcus* and *Lactobacillus*, which dominate the baby gut microbiota (Supplementary Table 5 and Supplementary Fig. 3), are deficient in these genes (Supplementary Fig. 16). By contrast, several of these early gut colonizers contain ECs involved in folate biosynthesis and metabolism (Supplementary Fig. 16). The conventional view of the developing infant gut is that the main change is in the representation of Bifidobacteria. Although differences in the representation of Bifidobacteria contribute to this effect, differences in vitamin metabolism among the rest of the bacteria remain even when all Bifidobacteria reads were excluded (data not shown). These changes in vitamin biosynthetic pathway representation in the microbiome correlate with published reports indicating that blood levels of folate decrease and cobalamin increase with age[22].

Besides cobalamin and folate, the relative abundance of ECs involved in the biosynthesis of vitamins B7 (biotin) (biotin synthase, EC2.8.16) and B1 (thiamine) (thiamine-phosphate diphosphorylase, EC2.5.1.3) are significantly higher in adult microbiomes than the microbiomes of babies (Supplementary Fig. 17 and Supplementary Table 7). Together, these findings suggest that the microbiota should be considered when assessing the nutritional needs of humans at various stages of development.

Random Forests analysis asks a different statistical question from ShotgunFunctionalizeR: that is, which genes or species are most discriminatory among different class labels, rather than which are most over/underrepresented, and tends to identify fewer features than ShotgunFunctionalizeR when applied to the same data. Random Forests analysis yielded 107 ECs that best discriminate the adult and baby microbiomes (Supplementary Table 7). These predictive ECs were among the most significantly different ECs determined by ShotgunFunctionalizeR and included ECs involved in the metabolism of cobalamin and folate (Supplementary Table 7). In addition, Random Forests showed that ECs involved in fermentation, methanogenesis and the metabolism of arginine, glutamate, aspartate and lysine were higher in the adult microbiomes, whereas ECs involved in the metabolism of cysteine and fermentation pathways found in lactic acid bacteria (acetolactate decarboxylase (EC4.1.1.5) and 6-phosphogluconate dehydrogenase (EC1.1.1.44)) were mainly represented in baby microbiomes (Supplementary Fig. 17).

Comparison of the representation of KEGG KOs between baby and adult microbiomes yielded essentially the same results as those reported with ECs. The only new finding was the overrepresentation of KEGG KOs assigned to a wide variety of ATP-binding cassette (ABC) transporters in baby microbiomes (Supplementary Table 7b).

Population- and age-specific differences

ShotgunFunctionalizeR, Random Forests and Spearman rank correlation analyses were all used to compare EC representation in fecal microbiomes as a function of predefined categories of geographic location and age. A total of 476 ECs were identified as being significantly different in US versus Malawian and Amerindian breastfed babies ($P < 0.0001$, ShotgunFunctionalizeR; Supplementary Table 8). The most prominent differences involved pathways related to vitamin biosynthesis and carbohydrate metabolism. Malawian and Amerindian babies had higher representation of ECs that were components of the vitamin B2 (riboflavin) biosynthetic pathway (Fig. 3a and Supplementary Fig. 18). These differences were not evident in adults (Supplementary Table 7). Riboflavin is found in human milk and in meat and dairy products. We did not measure the levels of these vitamins in mothers and in their breastmilk in the sampled populations, although it is tempting to speculate that the observed differences in baby microbiomes may represent an adaptive response to vitamin availability.

Studies in gnotobiotic mouse models indicate that the ability of members of the microbiota to access host-derived glycans plays a key part in establishing a gut microbial community[23,24]. As expected[4,5], compared with adults, baby microbiomes were enriched in ECs involved in the foraging of glycans represented in breastmilk and the intestinal mucosa (mannans, sialylated glycans, galactose and fucosyloligosaccharides; Supplementary Table 7). Several genes involved in using these host glycans are significantly overrepresented in Amerindian and Malawian baby microbiomes compared with US

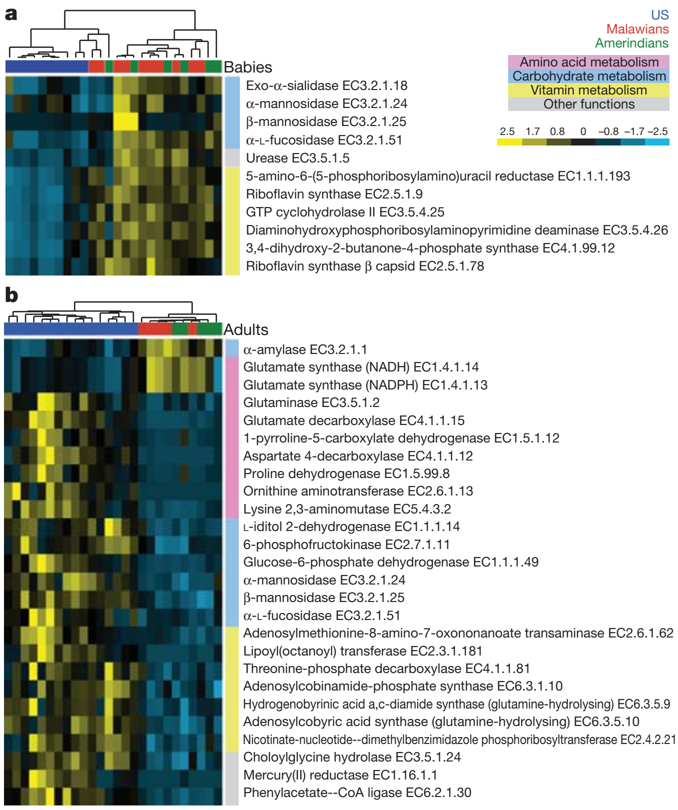


Figure 3 | Differences in the functional profiles of fecal microbiomes in the three study populations. Examples of KEGG ECs that showed the largest differences, as determined by Random Forests and ShotgunFunctionalizeR analyses, in proportional representation between US and Malawian/Amerindian populations. Shown are the relative abundances of genes encoding the indicated ECs (normalized by Z-score across all data sets). **a**, UPGMA (unweighted pair group method with arithmetic mean) clustering of 10 US, 10 Malawian and 6 Amerindian baby fecal microbiomes. **b**, UPGMA clustering of 16 US, 5 Malawian and 5 Amerindian adult fecal microbiomes.

baby microbiomes, most notably exo-α-sialidase and α-L-fucosidase (Fig. 3a and Supplementary Table 8). These population-specific biomarkers may reflect differences in the glycan content of breastmilk. In fact, the representation of these glycoside hydrolases decreases as Malawian and Amerindian babies mature and transition to a diet dominated by maize, cassava and other plant-derived polysaccharides. By contrast, α-fucosidase gene representation in the US infants increases with age and as they become exposed to diets rich in readily absorbed sugars (Supplementary Fig. 19d and Supplementary Table 9).

Another biomarker that distinguishes microbiomes based on age and geography is urease (EC3.5.1.5). Urease gene representation is significantly higher in Malawian and Amerindian baby microbiomes and decreases with age in these two populations, unlike in the United States, where it remains low from infancy to adulthood (Fig. 3a and Supplementary Fig. 19e). Urea comprises up to 15% of the nitrogen present in human breastmilk[25]. Urease releases ammonia that can be used for microbial biosynthesis of essential and nonessential amino acids[26,27]. Furthermore, urease has a crucial involvement in nitrogen recycling, particularly when diets are deficient in protein[28,29]. Under conditions in which dietary nitrogen is limiting, the ability of the microbiome to use urea would presumably be advantageous to both the microbial community and host. Although urease is generally attributed to *Helicobacter* and *Proteus* spp., the relative abundance of members of these two genera was low (<0.05%) and not significantly different between the three populations. Urease activity has been characterized previously in *Streptococcus thermophilus*[30]. Our analysis of shotgun reads that matched to the 126 reference gut genomes showed that the representation of five species that possess EC3.5.1.5 (*Bacteroides cellulosilyticus* WH2, *Coprococcus comes*, *Roseburia intestinalis*, *Streptococcus infantarius* and *S. thermophilus*) was significantly higher in Malawian and Amerindian baby microbiomes than in US baby microbiomes (Supplementary Table 5).

Further support of the role of diet in shaping the infant gut microbiome comes from the differences detected between breastfed and formula-fed babies who were part of the US infant twin cohort (see Supplementary Results and Supplementary Figs 2c, 8 and 20).

Differences in adult fecal microbiomes

Annotation of the shotgun sequencing data sets yielded a total of 1,349 ECs in the 26 adults surveyed: ShotgunFunctionalizeR showed that the representation of genes encoding 893 of these ECs was significantly different in US versus Malawian/Amerindian fecal microbiomes ($P < 0.005$ after multiple comparison correction; 433 overrepresented in US samples). By contrast, at this threshold only 445 ECs were identified as different between Malawian and Amerindian adults (see Supplementary Table 10 for a complete list). The Random Forests classifier revealed 52 ECs that were best at discriminating US versus non-US adult fecal microbiomes. These ECs were also identified by ShotgunFunctionalizeR as the most significantly different (Supplementary Table 10).

A typical US diet is rich in protein, whereas diets in Malawi and Amerindian populations are dominated by corn and cassava (Supplementary Table 1). The differences between US and Malawian/Amerindian microbiomes can be related to these differences in diet. The ECs that were the most significantly enriched in US fecal microbiomes parallel differences observed in carnivorous versus herbivorous mammals[31]. ECs encoding glutamate synthase have higher proportional representation in Malawian and Amerindian adult microbiomes and are also higher in herbivorous mammalian microbiomes[31] (Fig. 3b), whereas the degradation of glutamine was overrepresented in US as well as carnivorous mammalian microbiomes. Several ECs involved in the degradation of other amino acids were overrepresented in adult US fecal microbiomes: aspartate (EC4.1.1.12), proline (EC1.5.99.8), ornithine (EC2.6.1.13) and lysine (EC5.4.3.2) (Fig. 3b), as were ECs involved in the catabolism of simple sugars (glucose-6-phosphate dehydrogenase and 6-phosphofructokinase), sugar substitutes (L-iditol 2-dehydrogenase, which degrades sorbitol), and host glycans (α-mannosidase, β-mannosidase and α-fucosidase; Fig. 3b). By contrast, α-amylase (EC3.2.1.1), which participates in the degradation of starch, was overrepresented in the Malawian and Amerindian microbiomes, reflecting their corn-rich diet.

US microbiomes also had significant overrepresentation of ECs involved in vitamin biosynthesis (cobalamin (Fig. 3b and Supplementary Fig. 15), biotin and lipoic acid (Fig. 3b)), in the metabolism of xenobiotics (phenylacetate CoA ligase (EC6.2.1.30), which participates in the metabolism of aromatic compounds, and mercury reductase (EC1.16.1.1)), and in bile salt metabolism (choloylglycine hydrolase (EC3.5.1.24), perhaps reflecting a diet richer in fats (Fig. 3b)).

Effects of kinship on the microbiome across countries

Differences in social structures may influence the extent of vertical transmission of the microbiota and the flow of microbes and microbial genes among members of a household. Differences in cultural traditions also affect food, exposure to pets and livestock, and many other factors that could influence how and from where a gut microbiota/microbiome is acquired. We previously observed that adult monozygotic twins are no more similar to one another in terms of their gut bacterial community structure than are adult dizygotic twins[32]. This result suggests that the overall heritability of the microbiome is low. We confirmed that the phylogenetic architecture of the fecal microbiota of monozygotic Malawian co-twins ≤3 years of age is

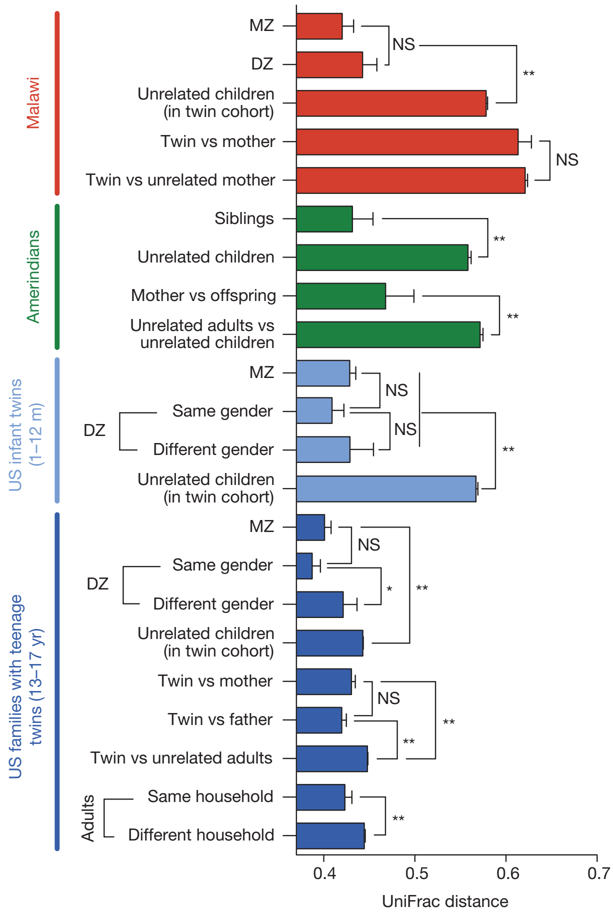


Figure 4 | Differences in the fecal microbiota between family members across the three populations studied. UniFrac distances between the fecal bacterial communities of family members were calculated ($n = 19$ Amerindian, 34 Malawian families, and 54 US families with teenage twins). DZ, dizygotic; MZ, monozygotic. Mean and s.e.m. values are plotted. The UniFrac matrix was permuted 1,000 times; P values represent the fraction of times permuted differences were greater than real differences. m, months; NS (not significant; $P > 0.05$), $^*P < 0.05$, $^{**}P < 0.005$.

no more similar than the microbiota of similarly aged dizygotic co-twins ($n = 15$ monozygotic and 6 dizygotic twin pairs). We found that this is also true for monozygotic and dizygotic twin pairs aged 1–12 months ($n = 16$ twin pairs), as well as teenaged twins (13–17 years old; $n = 50$ pairs) living together in the United States (Fig. 4).

Although biological mothers are in a unique position to transmit an initial inoculum of microbes to their infants during and after birth, our analysis of mothers of teenage US twins showed that their fecal microbiota were no more similar to their children than were those of biological fathers, and that genetically unrelated but co-habiting mothers and fathers were significantly more similar to one another microbially than were members of different families (Fig. 4; note that no fathers were sampled in Malawi and only four fathers in the Amerindian cohort). These latter observations emphasize the importance of a history of numerous common environmental exposures in shaping gut microbial ecology. Moreover, the similarity in the overall pattern of the effects of kinship on microbial community structure suggests that despite the large influence of cultural factors on which microbes are present in both children and adults in each population, the bases for the degree of similarity among members of a family are consistent across the three populations studied.

Prospectus

Our results emphasize that it is essential to sample a broad population of healthy humans over time, both in terms of their age, geography and cultural traditions, to discover features of gut microbiomes that are unique to different locations and lifestyles. In addition, we need to understand how the pressures of westernization are changing the microbial parts of our genetic landscape—changes that potentially mediate the suite of pathophysiological states correlated with westernization. Finally, given the need for global policies about sustainable agriculture and improved nutrition, it will be important to understand how we can match these policies not only to our varying cultural traditions but also to our varied gut microbiomes.

METHODS SUMMARY

Sample collection. Subjects were recruited for the present study using procedures approved by Human Studies Committees from Washington University in St Louis, Children's Hospital of Philadelphia, the University of Colorado, Boulder, the University of Malawi, the University of Puerto Rico, and the Venezuelan Institute for Scientific Research. Each fecal sample was frozen within 30 min of donation.

Multiplex DNA sequencing. Extracted genomic DNA was subjected to multiplex Illumina sequencing of the V4 region of bacterial 16S rRNA genes, as well as multiplex 454 pyrosequencing of total community DNA. See Methods for further details about the analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 February 2011; accepted 20 March 2012.
Published online 9 May 2012.

1. Mueller, S. *et al.* Differences in fecal microbiota in different European study populations in relation to age, gender, and country: a cross-sectional study. *Appl. Environ. Microbiol.* **72,** 1027–1033 (2006).
2. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
3. Li, M. *et al.* Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl Acad. Sci. USA* **105,** 2117–2122 (2008).
4. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14,** 169–181 (2007).
5. Koenig, J. E. *et al.* Microbes and Health Sackler Colloquium: Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4578–4585 (2011).
6. Favier, C. F., Vaughan, E. E., De Vos, W. M. & Akkermans, A. D. Molecular monitoring of succession of bacterial communities in human neonates. *Appl. Environ. Microbiol.* **68,** 219–226 (2002).
7. Tannock, G. W. What immunologists should know about bacterial communities of the human bowel. *Semin. Immunol.* **19,** 94–105 (2007).
8. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* **107,** 11971–11975 (2010).
9. Blaser, M. J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nature Rev. Microbiol.* **7,** 887–894 (2009).
10. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
11. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107,** 14691–14696 (2010).
12. Peach, S., Fernandez, F., Johnson, K. & Drasar, B. S. The non-sporing anaerobic bacteria in human faeces. *J. Med. Microbiol.* **7,** 213–221 (1974).
13. Mackie, R. I., Sghir, A. & Gaskins, H. R. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am. J. Clin. Nutr.* **69,** 1035S–1045S (1999).
14. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5,** e177 (2007).
15. Penders, J. *et al.* Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118,** 511–521 (2006).
16. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
17. Knights, D., Costello, E. K. & Knight, R. Supervised classification of human microbiota. *FEMS Microbiol. Rev.* **35,** 343–359 (2011).
18. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
19. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).
20. Kristiansson, E., Hugenholtz, P. & Dalevi, D. ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* **25,** 2737–2738 (2009).
21. Kräutler, B. Vitamin B_{12}: chemistry and biochemistry. *Biochem. Soc. Trans.* **33,** 806–810 (2005).
22. Monsen, A. L., Refsum, H., Markestad, T. & Ueland, P. M. Cobalamin status and its biochemical markers methylmalonic acid and homocysteine in different age groups from 4 days to 19 years. *Clin. Chem.* **49,** 2067–2075 (2003).
23. Martens, E. C., Chiang, H. C. & Gordon, J. I. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4,** 447–457 (2008).
24. Hooper, L. V., Xu, J., Falk, P. G., Midtvedt, T. & Gordon, J. I. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc. Natl Acad. Sci. USA* **96,** 9833–9838 (1999).
25. Harzer, G., Franzke, V. & Bindels, J. G. Human milk nonprotein nitrogen components: changing patterns of free amino acids and urea in the course of early lactation. *Am. J. Clin. Nutr.* **40,** 303–309 (1984).
26. Metges, C. C. *et al.* Incorporation of urea and ammonia nitrogen into ileal and fecal microbial proteins and plasma free amino acids in normal men and ileostomates. *Am. J. Clin. Nutr.* **70,** 1046–1058 (1999).
27. Millward, D. J. *et al.* The transfer of 15N from urea to lysine in the human infant. *Br. J. Nutr.* **83,** 505–512 (2000).
28. Meakins, T. S. & Jackson, A. A. Salvage of exogenous urea nitrogen enhances nitrogen balance in normal men consuming marginally inadequate protein diets. *Clin. Sci. (Lond.)* **90,** 215–225 (1996).
29. Langran, M., Moran, B. J., Murphy, J. L. & Jackson, A. A. Adaptation to a diet low in protein: effect of complex carbohydrate upon urea kinetics in normal man. *Clin. Sci. (Lond.)* **82,** 191–198 (1992).
30. Mora, D. *et al.* Characterization of urease genes cluster of *Streptococcus thermophilus*. *J. Appl. Microbiol.* **96,** 209–219 (2004).
31. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332,** 970–974 (2011).
32. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Wagoner and J. Manchester for superb technical assistance, plus B. Muegge, A. Grimm, A. Hsiao, N. Griffin and P. Tarr for suggestions, and M. Ndao, T. Tinnin and R. Mkakosya for patient recruitment and/or technical assistance. This work was supported in part by grants from the National Institutes of Health (DK078669, T32-HD049338), St. Louis Children's Discovery Institute (MD112009-201), the Howard Hughes Medical Institute, the Crohn's and Colitis Foundation of America, and the Bill and Melinda Gates Foundation. Parts of this work used the Janus supercomputer, which is supported by National Science Foundation grant CNS-0821794, the University of Colorado, Boulder, the University of Colorado, Denver, and the National Center for Atmospheric Research.

Author Contributions T.Y., R.K. and J.I.G. designed the experiments, M.J.M., I.T., M.G.D.-B., M.C., M.M., G.H., A.C.H., A.P.A., R.K., R.N.B., C.A.L., C.L. and B.W. participated in patient recruitment, T.Y. generated the data, T.Y., F.E.R., J.R., J.K., J.G.C., J.C.C., D.K., R.K. and J.I.G. analysed the results, T.Y., R.K. and J.I.G. wrote the paper.

Author Information DNA sequences have been deposited in MG-RAST (http://metagenomics.anl.gov/) under accession numbers 'qiime:850' for Illumina V4 16S rRNA data sets, and 'qiime:621' for fecal microbiome shotgun sequencing data sets. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.I.G. (jgordon@wustl.edu).

METHODS

Isolation of fecal DNA and multiplex sequencing. Each participant provided a fecal specimen that was frozen within 30 min. All samples were stored at −80 °C and subjected to a common protocol for DNA extraction. Fecal samples were pulverized with a mortar and pestle at −80 °C. Genomic DNA was extracted from 400 mg aliquots of frozen pulverized samples. Methods for multiplex Illumina sequencing of V4 amplicons have been described[33].

For multiplex shotgun 454 Titanium FLX pyrosequencing, each fecal community DNA sample was randomly fragmented by nebulization (500–800 base pairs) and then labelled with a distinct Multiplex IDentifier (MID; Roche) according the manufacturer's protocol (Rapid Library preparation for FLX Titanium, Roche). Equivalent amounts of 12 MID-labelled samples were pooled before each pyrosequencer run.

Data analysis. 16S rRNA OTUs were picked from the Illumina reads using a closed-reference OTU picking protocol against the Greengenes database clustered at 97% identity (that is, uclust_ref: sequences are clustered against a reference database, and reads that do not match the database are excluded from further analyses) with uclust using the QIIME suite of software tools[10] version 1.3.0-dev (pick_otus.py parameters: –max_accepts 1 –max_rejects 8 –stepwords 8 –word_length 8). Of the 1,093,740,274 Illumina reads from the V4 region of bacterial 16S rRNA genes that passed the QIIME quality filters, 87.1% (952,115,802) matched a reference sequence at ≥97% nucleotide sequence identity. Taxonomy assignments were associated with OTUs based on the taxonomy associated with the Greengenes reference sequence defining each OTU. UniFrac distances between samples were calculated using the Greengenes reference tree. Greengenes reference sequences, trees and taxonomy data used in this analysis can be found at: http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/Reference_OTUs_for_Pipelines/Caporaso_Reference_OTUs/gg_otus_4feb2011.tgz.

A table of OTU counts per sample was generated and used in combination with the tree to calculate α and β diversity. To generate unweighted UniFrac distance matrices, all communities were rarefied to 290,609 V4 16S rRNA reads per sample. Unweighted UniFrac rather than weighted UniFrac was used for analyses owing to the large differences in taxonomic representation among the samples. Nonetheless, the patterns were similar with weighted UniFrac (data not shown). Rarefaction analysis was conducted using the QIIME scripts multiple_rarefaction.py, alpha_diversity.py and collate_alpha.py. The QIIME metric 'observed species' was used to estimate α diversity in the data set.

Clustering analysis. Testing for discrete clusters was performed on the rarefied versions of the 16S rRNA OTU relative abundance tables. OTU counts were binned into genus-level taxonomic groups according to the taxonomic assignments described earlier. Several distance measures were considered, including Jensen–Shannon divergence, Bray–Curtis and weighted/unweighted UniFrac distances. Clustering was performed via partitioning around medoids in the R package cluster[34]. The choice of number of clusters and quality of the resulting clusters were assessed by maximizing the silhouette index[35]. Traditionally, silhouette indices of 0.5 or above have been considered evidence of reasonable clustering structure. Although some silhouette scores above 0.5 were found in this data set (for example, for two clusters when clustering all adult populations with Jensen–Shannon divergence), reclustering within different subpopulations (for example, individual countries) introduced new cluster boundaries with silhouette scores still near or above 0.5, indicating that silhouette index scores may need to be substantially above 0.5 to claim clustering structure for microbial enterotype testing. We also tested for discrete clusters using the prediction strength measure[36], which showed negative results for all distances measures but unweighted UniFrac (prediction strength = 0.963 ± 0.012 (mean ± s.d.)). We estimated the s.d. by tenfold jackknifing.

Shotgun sequences from fecal microbiomes. Shotgun reads were filtered using custom Perl scripts and publicly available software to remove (1) all reads <60 nucleotides; (2) Titanium pyrosequencing reads with two continuous and/or three total degenerate bases (N); (3) all duplicates (a known artefact of pyrosequencing), defined as sequences in which the initial 20 nucleotides are identical and that share an overall identity of >97% throughout the length of the shortest read[37]; and (4) all sequences with significant similarity to human reference genomes (Blastn E-value threshold $\leq 10^{-5}$, bitscore ≥ 50, percentage identity ≥ 75%) to ensure the continued de-identification of samples.

Searches against the database of 126 human gut bacterial genomes were conducted with Blastn. A sequence read was annotated as the best hit in the database if the E-value was $\leq 10^{-5}$, the bit score was ≥50, and the alignment was at least 95% identical between query and subject. Relative abundances of reads mapped to each of the 126 genomes were adjusted to genome sizes. Searches against the protein-coding component of the KEGG database (v58) and against COG (v8.3) were conducted with BLASTX. (Note that when we performed searches against a separate KEGG database of intergenic regions alone, very few hits were observed.) Counts were normalized to the mapped reads. In total, 40 ± 8% reads were mapped to KEGG KOs and 56 ± 11% to COG; 44 ± 16% of the reads mapped to the 126 gut genomes using 95% sequence similarity cut-off. Unmapped reads were excluded from the analyses shown in the main text, although repeating the analyses including these reads had little effect on the results. To quantify the differences in KEGG EC profiles among fecal microbiomes, evenly rarefied matrices of EC counts were created with all samples, and Hellinger distances were calculated using QIIME.

Spearman rank correlations were carried out using the R statistical software[38]. To identify bacterial taxa that change with increasing age in each population, the proportion of reads that map to each of the 126 reference sequenced human gut genomes in each fecal microbiome was identified. The relative abundance of reads from each genome was then correlated with age (years) for each geographic region. To identify genes encoding ECs that change with age, the proportion of reads annotated with each EC in each fecal microbiome was identified. The relative abundance of each EC was subsequently correlated with age (years) for each geographic region.

Random Forests analysis. Random Forests analysis was applied as described in ref. 8, using the randomForest package in R[39] with 500 trees and all default settings. The generalization error was estimated using fivefold cross-validation for all comparisons involving adults from the 16S rRNA data; leave-one-out cross-validation was used for all other comparisons. For each comparison, the relevant subset of samples was extracted from the table of OTU or EC counts, and all singleton OTUs/ECs (or all OTUs/ECs present in fewer than 5 samples for the 16S rRNA comparisons involving adults) were subsequently removed. Random Forests analysis was performed for each comparison on 100 rarefied versions of the data, and the average cross-validation error estimates and OTU/EC importance estimates were reported. Rarefaction depths were chosen manually to include all samples without exceptionally low total sequences. The chosen depth for each comparison and the resulting number of samples are shown in Supplementary Tables 6–8 and Supplementary Fig. 6. For the analysis shown in Supplementary Fig. 6a, we compared the generalization errors obtained when using 16S rRNA-based OTUs from the Illumina V4 data sets at various sequencing depths. For direct evaluation of the predictive strength of the Illumina-based OTUs, we rarefied at the lowest observed depth of 305,631 sequences for each classification task, as well as at sequencing depths of 100, 1,000, 10,000, 100,000 and 1,000,000 reads per sample. The mean and s.d. of the cross-validation error were estimated for each classification tasking using ten independent rarefactions of the data. We also included the expected 'baseline' error obtained by a classifier that simply predicts the most common class label.

Data deposition. DNA sequences have been deposited in MG-RAST (http://metagenomics.anl.gov/) under accession numbers 'qiime:850' for Illumina V4 16S rRNA data sets, and 'qiime:621' for fecal microbiome shotgun sequencing data sets.

33. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12,** R50 (2011).
34. Kaufman, L. & Rousseeuw, P. J. *Finding Groups in Data: an Introduction to Cluster Analysis* Ch. 2 68–125 (Wiley, 1990).
35. Rousseeuw, P. J. Silhouettes — a graphical aid to the interpretation and validation of cluster-analysis. *J. Comput. Appl. Math.* **20,** 53–65 (1987).
36. Tibshirani, R. & Walther, G. Cluster validation by prediction strength. *J. Comput. Graph. Statist.* **14,** 511–528 (2005).
37. Teal, T. K. & Schmidt, T. M. Identifying and removing artificial replicates from 454 pyrosequencing data. *Cold Spring Harb. Protoc.* **2010,** pdb.prot5409 (2010).
38. R Development Core Team.. R: A Language and Envirnoment for Statistical Computing (R Foundation for Statistical Computing, 2010).
39. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2,** 18–22 (2002).

doi:10.1038/nature11162

NPR3 and NPR4 are receptors for the immune signal salicylic acid in plants

Zheng Qing Fu[1]*, Shunping Yan[1]*, Abdelaty Saleh[1]*, Wei Wang[1], James Ruble[2], Nodoka Oka[3], Rajinikanth Mohan[1], Steven H. Spoel[4], Yasuomi Tada[5], Ning Zheng[2] & Xinnian Dong[1]

Salicylic acid (SA) is a plant immune signal produced after pathogen challenge to induce systemic acquired resistance. It is the only major plant hormone for which the receptor has not been firmly identified. Systemic acquired resistance in *Arabidopsis* requires the transcription cofactor nonexpresser of PR genes 1 (NPR1), the degradation of which acts as a molecular switch. Here we show that the NPR1 paralogues NPR3 and NPR4 are SA receptors that bind SA with different affinities. NPR3 and NPR4 function as adaptors of the Cullin 3 ubiquitin E3 ligase to mediate NPR1 degradation in an SA-regulated manner. Accordingly, the *Arabidopsis npr3 npr4* double mutant accumulates higher levels of NPR1, and is insensitive to induction of systemic acquired resistance. Moreover, this mutant is defective in pathogen effector-triggered programmed cell death and immunity. Our study reveals the mechanism of SA perception in determining cell death and survival in response to pathogen challenge.

After pathogen challenge, host cells have to make a life-and-death decision to fend off infection. Recognition of a pathogen effector by a host resistance protein can lead to effector-triggered immunity (ETI), characterized by rapid programmed cell death (PCD) known as the hypersensitive response[1]. The clearly defined boundary of the hypersensitive response indicates the presence of a mechanism that controls cell death and survival. Despite intense studies of plant mutants defective in controlling the spread of PCD[2], the regulatory mechanism still remains a mystery.

Localized PCD can induce systemic acquired resistance (SAR) through the production of the immune signal, salicylic acid (SA)[3]. SA triggers global transcriptional reprogramming and resistance to a broad spectrum of pathogens. The receptor for SA has been sought after for many years, mainly through biochemical purification of SA-binding proteins[4–6]. However, genetic data for these SA-binding proteins, which include a catalase, a chloroplast carbonic anhydrase, and a methyl SA esterase, suggest that none of them functions as a bona fide SA receptor. By contrast, genetic studies of SA-insensitive mutants have strongly suggested that NPR1, which contains a BTB (bric à brac, tramtrack, broad-complex) domain, an ankyrin repeat domain and a nuclear localization sequence, is a potential SA receptor[7]. However, the NPR1 protein does not have considerable SA binding activity under different test conditions (Supplementary Fig. 2).

Instead of direct binding, SA has been shown to control the nuclear translocation of NPR1 through cellular redox changes[8]. In the absence of pathogen challenge, NPR1 is retained in the cytoplasm as an oligomer through redox-sensitive intermolecular disulphide bonds. After induction, these disulphide bonds are reduced, releasing NPR1 monomers into the nucleus, where NPR1 acts as a cofactor for transcription factors, such as TGAs, to induce defence-related genes. In the absence of a functional NPR1 protein, SA-induced transcriptional reprogramming is almost completely blocked.

The presence of a BTB domain in NPR1 suggests that, like other BTB domain-containing proteins, it may interact with Cullin 3 (CUL3) E3 ligase and mediate substrate degradation[9]. However, our research led to the surprising finding that the NPR1 protein itself is degraded by the proteasome. Although NPR1 is degraded in the nucleus of resting cells to dampen basal expression of defence genes, it is phosphorylated after immune activation at an IκB-like phosphodegron motif, ubiquitinylated by a CUL3 E3 ligase, and degraded to sustain maximum levels of target gene expression probably through accelerated recycling of the transcription initiation complex[10]. Blocking NPR1 degradation by mutating the IκB-like phophodegron in NPR1 or the two *CUL3* genes (*cul3a cul3b*) in *Arabidopsis* led to increased basal resistance, but insensitivity to SAR induction. Therefore, nuclear accumulation of NPR1 is needed for basal defence gene expression and resistance, whereas its subsequent turnover is required for establishing SAR.

NPR3 and NPR4 are CUL3 adaptors for NPR1 degradation

In a search for the adaptor proteins of the CUL3 E3 ligase that specifically target NPR1 for degradation, we considered its paralogues, NPR3 and NPR4, as possible candidates, because both contain the BTB domain as well as an extra protein–protein interaction domain (ankyrin repeat) (Supplementary Fig. 3), which are typical for CUL3 substrate adaptors[9]. More importantly, despite their sequence similarities to NPR1, the *npr3 npr4* double mutant has the opposite phenotype of *npr1* in that it exhibits enhanced disease resistance[11], a phenotype reminiscent of the *cul3a cul3b* mutant[10].

To test our hypothesis that NPR3 and NPR4 are CUL3 adaptors for NPR1 degradation, we examined the accumulation of NPR1 protein in wild-type, *npr3*, *npr4* and *npr3 npr4* mutant plants. NPR1 protein levels were higher in the *npr4* and *npr3 npr4* mutants than in the wild type in the absence of exogenous SA, and increased faster in the *npr3*, *npr4* and *npr3 npr4* mutants compared with wild type in response to SA treatment (Fig. 1a). The effects of *npr3* and *npr4* on NPR1 were probably post-transcriptional, as *NPR1* transcripts were not increased in these mutants (Supplementary Fig. 4). To prove our hypothesis

[1]Howard Hughes Medical Institute–Gordon and Betty Moore Foundation, Department of Biology, PO Box 90338, Duke University, Durham, North Carolina 27708, USA. [2]Howard Hughes Medical Institute, Department of Pharmacology, University of Washington, PO Box 357280, Seattle, Washington 98195, USA. [3]Faculty of Agriculture, Kagawa University, Miki, Kagawa 761-0795, Japan. [4]Institute of Molecular Plant Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK. [5]Life Science Research Center, Institute of Research Promotion, Kagawa University, 2393 Ikenobe, Miki-cho, Kita-gun, Kagawa 761-0795, Japan.
*These authors contributed equally to this work.

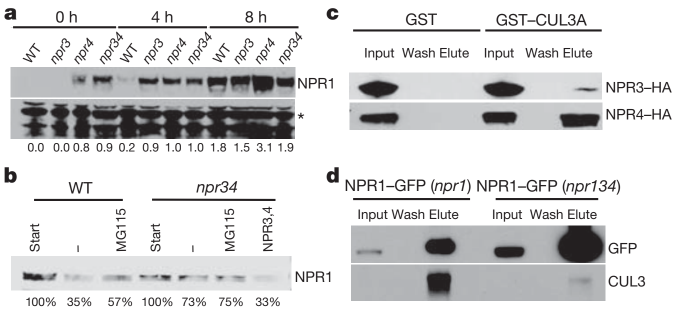


Figure 1 | NPR3 and NPR4 mediate degradation of NPR1. **a**, NPR1 protein levels in wild type (WT), *npr3*, *npr4* and *npr3 npr4* (*npr34*) plants treated with 0.5 mM SA. The NPR1 level (shown at bottom) was determined on the basis of the ratio of the NPR1 band intensity to that of the non-specific band (asterisk). **b**, GST–NPR1 degradation in extracts from wild-type or *npr3 npr4* double mutant (*npr34*), with or without (−) MG115 or with recombinant His–MBP–NPR3 and His–MBP–NPR4 proteins (NPR3,4). **c**, *In vitro* pull-down assay of GST–CUL3A and NPR3–HA and NPR4–HA. **d**, Co-immunoprecipitation of NPR1–GFP and CUL3 in *npr1* and *npr1 npr3 npr4* (*npr134*) plants.

further, we performed *in vitro* degradation experiments using purified recombinant glutathione *S*-transferase (GST)-tagged NPR1 protein. We found that after 15 min of incubation, the recombinant NPR1 protein was degraded in the wild-type plant extract, but not in *npr3 npr4* (Fig. 1b). The addition of purified recombinant NPR3 and NPR4 proteins tagged with histidine (His) and maltose binding protein (MBP) to the extract complemented the mutant phenotype, supporting a role of NPR3 and NPR4 in mediating NPR1 degradation. This degradation is probably through the proteasome, as application of the proteasome inhibitor MG115 stabilized the protein (Fig. 1b). To demonstrate further that NPR3 and NPR4 act as adaptors for the CUL3 E3 ligase, we first performed pull-down experiments using *in vitro* translated haemagglutinin (HA)-tagged NPR3 (NPR3–HA) and NPR4–HA. We found that CUL3A could pull down NPR3 and NPR4, with NPR4 showing a stronger interaction (Fig. 1c). Then we performed a co-immunoprecipitation assay using transgenic plants constitutively expressing NPR1–green fluorescent protein (GFP) in *npr1* and *npr1 npr3 npr4* mutants. We found that the amount of the endogenous CUL3 protein pulled down by NPR1–GFP was significantly reduced in the *npr1 npr3 npr4* triple mutant compared with the *npr1* single mutant (Fig. 1d), indicating that the NPR1–GFP interaction with CUL3 requires NPR3 and NPR4. These results further support our hypothesis that NPR4 and NPR3 are CUL3 adaptors for the degradation of NPR1 before and after SA induction, respectively (Fig. 1a).

SA affects NPR1–NPR3 and NPR1–NPR4 interactions

Proteasome-mediated protein degradation has a crucial role in regulating plant hormone receptors[12]. In some of these cases, the hormones act as a molecular glue to enable the formation of the receptor complex[13,14], which includes the substrate adaptor for the E3 ligase and the corresponding substrate. Our data show that proteasome-mediated degradation of NPR1 is also involved in SA signalling[10], although a different E3 ligase (CUL3, instead of CUL1) is used.

To test the possibility that SA is part of the NPR1–NPR3/4 complex, we performed a yeast two-hybrid (Y2H) assay. Using NPR3 as bait and NPR1 as prey, little growth was observed on plates without SA (Fig. 2a). However, yeast growth was observed on plates supplemented with 100 μM SA or with the functional analogue of SA 2,6-dichloroisonicotinic acid (INA)[15], but not on plates with 4-hydroxybenzoic acid (4-HBA)[6], which cannot induce SAR. Interestingly, although SA promoted the NPR1–NPR3 interaction, it disrupted the interaction between NPR1 and NPR4 (Fig. 2b). Moreover, NPR3 and NPR4 could form both homodimers and heterodimers with each other in the presence of SA and INA, but not 4-HBA (Fig. 2c). This suggests that NPR3 and NPR4 not only control NPR1 stability, but also self-regulate. Because NPR1 did not form homodimers with or without SA and interacted with NPR2 independently of SA (Supplementary Fig. 5), we focused on the regulatory roles of NPR3 and NPR4.

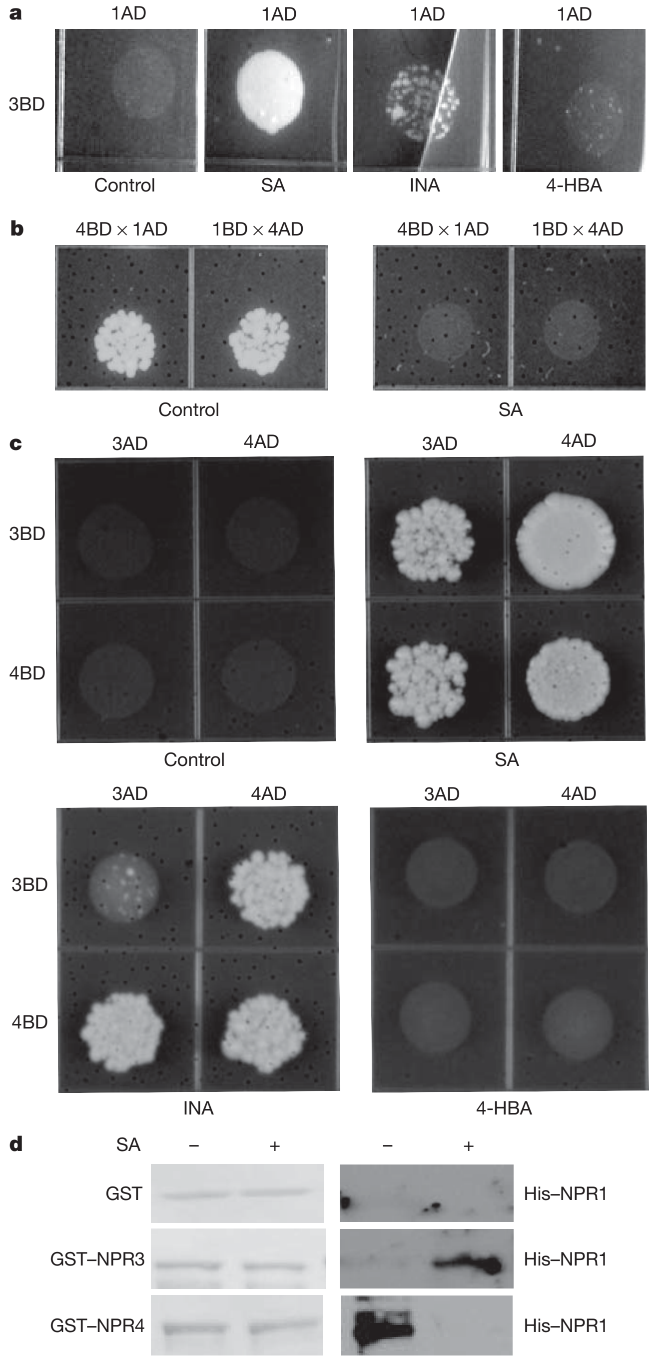


Figure 2 | SA directly regulates interactions between NPR proteins. **a**–**c**, Yeast two-hybrid assay to test interactions between NPR1 and NPR3 (**a**), NPR1 and NPR4 (**b**), and NPR3 and NPR4 (**c**). Diploid yeast cells were spotted on plates (SD media lacking Trp, Leu and His, plus 3 mM 3-aminotriazole) without (control) or with 100 μM SA, INA or 4-HBA. AD, activation domain; BD, DNA-binding domain. 1, NPR1; 3, NPR3; 4, NPR4. **d**, *In vitro* pull-down assays between His–MBP–NPR1 and GST–NPR3 and GST–NPR4 in the presence or absence of 100 μM SA.

To validate the Y2H data further, we performed *in vitro* pull-down assay. As shown in Fig. 2d, using the GST–NPR3 protein, we were able to pull down His–MBP–NPR1 only in the presence of SA. By contrast, GST–NPR4 could pull down His–MBP–NPR1 only in the absence of SA, indicating that the NPR1–NPR4 interaction was disrupted by SA.

NPR3 and NPR4 bind SA with different affinities

Both the Y2H and the *in vitro* pull-down results strongly suggest that SA directly binds to NPR3 and NPR4 to control their interactions with NPR1. To prove that NPR3 and NPR4 are SA receptors, we measured their SA-binding activities using [^{3}H]-SA. We found that both GST-tagged NPR3 and NPR4 recombinant proteins bound [^{3}H]-SA (Fig. 3a, b and Supplementary Fig. 2a). Next, we assessed whether active or inactive SA analogues could compete with [^{3}H]-SA to bind to GST–NPR3 and GST–NPR4. The active SAR inducer 5-chlorosalicylic acid[6] and INA reduced the binding of [^{3}H]-SA to GST–NPR3 and GST–NPR4, whereas 4-HBA had little effect (Fig. 3a, b). To assess the binding affinity of NPR3 and NPR4, we performed saturation binding experiments. Whereas NPR4 had a classical saturation curve (Fig. 3c), NPR3 binding could not be saturated even with 1,000 nM [^{3}H]-SA, indicating that NPR3 has a lower affinity than NPR4. Accordingly, the binding of SA to NPR3 was slower than NPR4 (Supplementary Fig. 6). Next, we analysed the saturation binding data with GraphPad Prism using different models, and found that the model 'one site—specific binding with Hill slope' is significantly better than the other models, which indicates that there are several binding sites or fractions in NPR3 and NPR4. The dissociation constant (K_d) for NPR4 was 46.2 ± 2.35 nM (mean ± s.e.m.) with a Hill coefficient (h) of 0.830 ± 0.0314. To check the cooperativity of different binding sites, we carried out dissociation experiments by the addition of 1 mM non-radioactive-labelled SA (cold SA) or by infinite dilution. The dissociation curves (Fig. 3d) indicate that NPR4 has several SA-binding sites, and the lack of overlap between the two curves suggests negative cooperativity between these binding sites (the first binding reduces the affinity for subsequent binding). The K_d value for NPR3 (981 nM, Supplementary Fig. 7) was significantly higher than 100 nM, which made saturation binding an inappropriate way to calculate the K_d. Therefore, we performed a homologous competitive binding assay (Fig. 3e). The half-maximum inhibitory concentration (IC_{50}) was calculated to be 1,811 nM ($\log(IC_{50}) = 3.26 \pm 0.0901$) with a Hill coefficient of 0.554 ± 0.0612. Through these analyses, we demonstrated that NPR3 and NPR4 bind SA specifically and with different affinities.

To examine the receptor complex further, we performed gel filtration analysis on the purified recombinant NPR4 protein—the receptor with the higher affinity to SA. Because the recombinant NPR4 protein spontaneously oligomerized *in vitro* in the absence of a reducing agent, our analysis focused on samples pretreated with 100 mM dithiothreitol (DTT) followed by dialysis against 5 mM DTT. We discovered that NPR4 was present in an estimated tetrameric form, which was competent in binding to SA (Fig. 3f). Notably, SA binding did not change the gel filtration elution profile of the protein. Further experiments are required to investigate how SA affects the receptor complexes to make them either more accessible (that is, NPR3 binding to NPR1) or less accessible (that is, NPR4 binding to NPR1) for substrate binding.

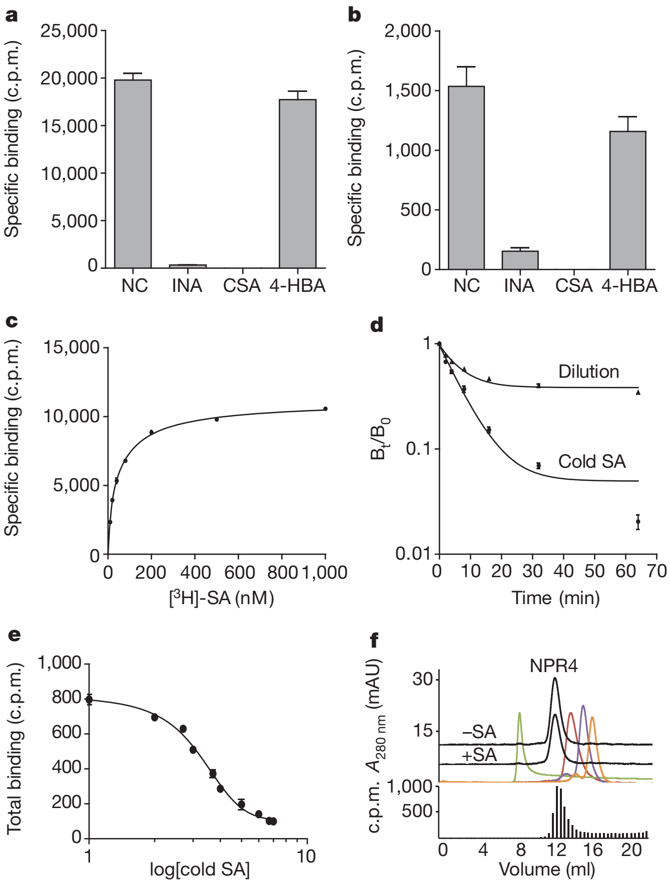


Figure 3 | NPR3 and NPR4 bind SA. **a**, **b**, Competition binding assay of NPR4 (**a**) and NPR3 (**b**). c.p.m., counts per minute; CSA, 5-chlorosalicylic acid; NC, no competitor. **c**, Saturation binding assay of NPR4. $K_d = 46.2 \pm 2.35$ nM, $h = 0.830 \pm 0.0314$. **d**, Dissociation assay of NPR4. The dissociation was initiated by the addition of 1 mM non-radioactive-labelled SA (cold SA) or by infinite dilution. B_0 and B_t are total binding before and after dissociation, respectively. **e**, Homologous competitive binding assay of NPR3. $IC_{50} = 1{,}811$ nM ($\log(IC_{50}) = 3.26 \pm 0.0901$), $h = 0.554 \pm 0.0612$. **f**, Size exclusion chromatography showing that NPR4 tetramer binds SA (black). Top panel, elution profile. Green, red, purple and orange peaks correspond to 2,000, 158, 75 and 44 kDa, respectively. Bottom panel, total binding of [^{3}H]-SA in different fractions. mAU, milli-absorbance units. Error bars represent s.d. (n = 2 or 3).

The *npr3 npr4* double mutant is defective in SAR and ETI

To understand the biological significance of NPR3/4-mediated degradation of NPR1, a positive regulator of SAR, we first performed SAR tests in the *npr3*, *npr4* and *npr3 npr4* mutants using *Pseudomonas syringae* pv. *maculicola* ES4326 (*Psm* ES4326). Consistent with a previous report[11], there was a significant reduction in *Psm* ES4326 growth in the *npr3 npr4* double mutant without SAR induction (Fig. 4a). However, even after SAR induction by local inoculation of avirulent *Psm* ES4326/*avrRpt2*, no further reduction in growth of virulent *Psm* ES4326 in systemic tissue was observed in the *npr3 npr4* double mutant. To a lesser degree, SAR was also defective in the *npr3* single mutant. Thus, stabilization of NPR1 protein in the *npr3* and *npr3 npr4* mutants rendered these plants insensitive to SAR induction. This phenotype is similar to that observed in the *cul3a cul3b* double mutant[10], validating the role of NPR3 and NPR4 in CUL3-mediated degradation of NPR1 and SAR.

On the basis of our knowledge that SAR and ETI are two distinct defence strategies, with the former promoting cell survival and the latter triggering PCD, we then tested the *npr* mutants for ETI using *Pseudomonas* strains expressing different effectors. Surprisingly, we found that the *npr3 npr4* mutant failed to undergo PCD (Fig. 4b and Supplementary Fig. 8a) as quantified by ion leakage (Fig. 4c), and was compromised in resistance triggered by the effectors (Fig. 4d). The same phenotypes were observed in different mutant alleles of *npr3*, *npr4* and *npr3 npr4* (Supplementary Fig. 8b). The ETI deficiency observed in *npr3 npr4* is probably caused by the increased accumulation of NPR1, because this phenotype was suppressed in the *npr1 npr3 npr4* triple mutant.

To observe NPR1 turnover in response to pathogen challenge *in situ*, we inoculated *Psm* ES4326/*avrRpt2* in the *35S:NPR1(C82A)-GFP* transgenic plant, in which NPR1 is constitutively localized in the nucleus (Fig. 4e)[16]. Eleven hours after inoculation, some cells showed increased chlorophyll leakage (Fig. 4f, red) with overlapping accumulation of phenolic compounds (Fig. 4e, larger green spots) indicative of PCD (Fig. 4g, yellow), whereas other cells were still intact. The NPR1(C82A)–GFP fluorescence was markedly reduced inside the inoculated region (Fig. 4g, h). Notably, the NPR1(C82A)–GFP fluorescence level was the highest in the cells surrounding the hypersensitive

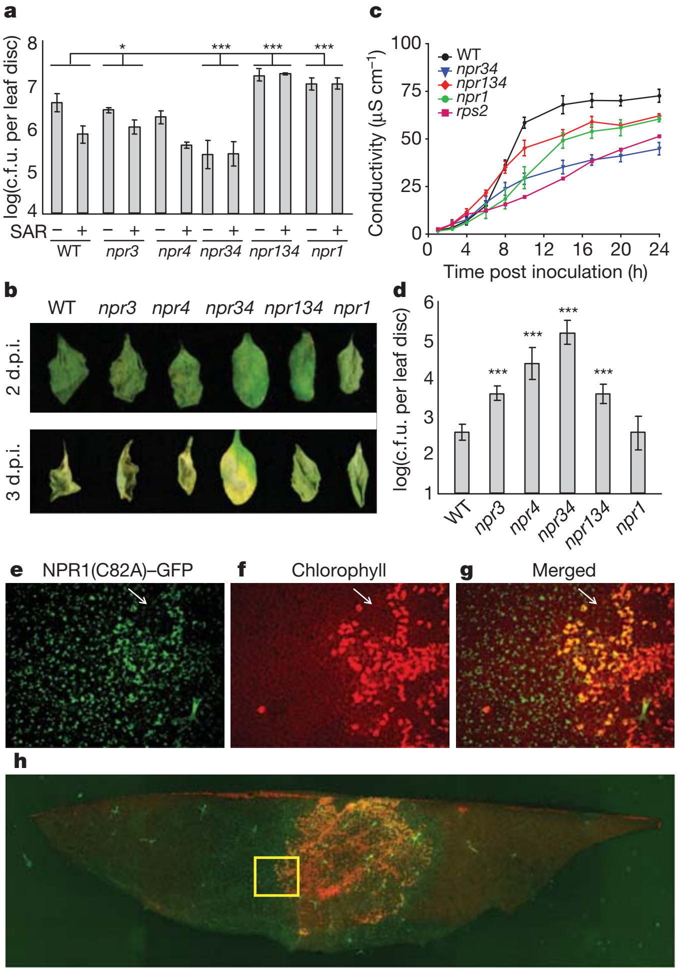


Figure 4 | SA receptors control NPR1 stability to regulate SAR and ETI. **a**, SAR test in wild-type, *npr3*, *npr4*, *npr3 npr4* (*npr34*), *npr1 npr3 npr4* (*npr134*) and *npr1*. c.f.u., colony forming units. **b**–**d**, ETI test in different mutants using *Psm* ES4326/*avrRpt2*. **b**, The hypersensitive response phenotype, 2 and 3 days post inoculation (d.p.i.). **c**, Ion leakage measurement. *rps2*, an *avrRpt2*-insensitive mutant. Error bars represent s.d., $n = 4$. **d**, Growth of *Psm* ES4326/*avrRpt2*. Error bars in **a** and **d** represent 95% confidence intervals, $n = 6$–8. $*P < 0.05$, $***P < 0.001$. **e**–**g**, Close-up images of an infection site by *Psm* ES4326/*avrRpt2*. Arrows point to intact cells inside the inoculated area. **h**, Image of the whole infection site showing high NPR1(C82A)–GFP accumulation surrounding the PCD zone. The rectangle shows the area from which the close-up images in **e**–**g** were taken. Yellow staining in **g** and **h** indicates dead cells, green indicates NPR1(C82A)–GFP. Original magnification, ×10 (**e**–**g**) and ×2 (**h**).

response lesion (Fig. 4h), consistent with the genetic data suggesting that NPR1 is an inhibitor of PCD during ETI.

Discussion

Through this study, we identified the NPR1 paralogues NPR3 and NPR4 as receptors for the immune signal SA. These receptors have different binding affinities to SA (Fig. 3), suggesting that they may be differentially responsive to spatiotemporal changes in cellular SA concentrations. SA controls accessibility of the CUL3 ligase adaptors NPR3 and NPR4 to their substrate NPR1 (Fig. 2), thereby regulating NPR1 stability and activity (Figs 1 and 4).

On the basis of our findings, we present a working model for the regulation of NPR1 by NPR3 and NPR4 in response to different SA levels (Supplementary Fig. 1). In the absence of pathogen challenge, NPR4 constantly removes most of the NPR1 protein by CUL3-NPR4-mediated degradation. This degradation is important to prevent spurious activation of resistance. However, basal SA is required to disrupt some of the NPR1–NPR4 interactions to maintain the basal level of NPR1. This is crucial because SA-deficient plants, *eds5* (ref. 17), *ics1* (also known as *eds16*)[18] and the *nahG* transgenic line expressing an SA-degrading enzyme[19], are impaired in maintaining NPR1 homeostasis (Supplementary Fig. 9), resulting in enhanced disease susceptibility (Supplementary Fig. 1a, b). After challenge by pathogens that trigger ETI, SA levels increase both locally and systemically[20] to form a concentration gradient from the infection site[21,22]. Previous research has shown that high levels of SA facilitate PCD[23,24], and the spread of PCD may be controlled by the activities of proteins such as LSD1 (a zinc-finger protein) and Atrboh D (an NADPH oxidase). Because the *npr3 npr4* double mutant can no longer undergo PCD in response to pathogen effectors, we propose that NPR1, which over-accumulates in the *npr3 npr4* mutant, can act as a negative regulator of PCD. Our finding is in line with a previous report suggesting that NPR1 suppresses hypersensitive response[25]. In support of this function of NPR1, the NPR1–GFP signal is the lowest inside the developing hypersensitive response (Fig. 4e–h) owing to CUL3-NPR3-mediated degradation of NPR1 (Supplementary Fig. 1c). In neighbouring cells, the lower SA level limits NPR1–NPR3 interaction, enabling NPR1 to accumulate in the margin of the hypersensitive response to restrict the spread of PCD and establish SAR (Supplementary Fig. 1d).

METHODS SUMMARY

Protein analysis was carried out as previously described[10]. Total NPR1 protein was detected by an antibody against NPR1. For the *in vitro* degradation assay, the purified recombinant GST–NPR1 was incubated with plant extracts and detected by an anti-GST antibody (GE Healthcare). The *in vitro* pull-down assays for CUL3A, NPR3 and NPR4 were performed using purified recombinant GST–CUL3A and *in vitro* translated NPR3–HA and NPR4–HA, which were detected by an anti-HA antibody (GenScript). Purified recombinant GST–NPR3 and GST–NPR4 proteins retained on the glutathione agarose beads were used to pull down purified His–MBP–NPR1. Bound His–MBP–NPR1 was detected by western blot using an anti-His antibody (GenScript). For co-immunoprecipitation, the immunoprecipitation was performed using an anti-GFP antibody (Abcam) and the western blot using an anti-CUL3A antibody[26] and an anti-GFP antibody (Clontech). Y2H assays were carried out using the Matchmaker system (Clontech). The interactions were determined by yeast growth on selective medium (SD media lacking Trp, Leu and His, plus 3 mM 3-aminotriazole) with or without 100 μM SA, INA or 4-HBA. The SA-binding assays were performed as described[6] with modifications using purified recombinant GST–NPR3 or GST–NPR4 and [^{3}H]-SA (American Radiolabelled Chemicals). Pathogen infection and ion leakage assay using *Psm* ES4326 with or without *avrRpt2* were carried out as previously described[10,27].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 November 2011; accepted 26 April 2012.
Published online 16 May 2012.

1. Jones, J. D. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).
2. Lorrain, S., Vailleau, F., Balague, C. & Roby, D. Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants? *Trends Plant Sci.* **8,** 263–271 (2003).
3. Durrant, W. E. & Dong, X. Systemic acquired resistance. *Annu. Rev. Phytopathol.* **42,** 185–209 (2004).
4. Chen, Z., Silva, H. & Klessig, D. Involvement of reactive oxygen species in the induction of sytemic acquired resistance by salicylic acid in plants. *Science* **262,** 1883–1886 (1993).
5. Park, S. W., Kaimoyo, E., Kumar, D., Mosher, S. & Klessig, D. F. Methyl salicylate is a critical mobile signal for plant systemic acquired resistance. *Science* **318,** 113–116 (2007).
6. Slaymaker, D. H. *et al.* The tobacco salicylic acid-binding protein 3 (SABP3) is the chloroplast carbonic anhydrase, which exhibits antioxidant activity and plays a role in the hypersensitive defense response. *Proc. Natl Acad. Sci. USA* **99,** 11640–11645 (2002).
7. Cao, H., Glazebrook, J., Clark, J. D., Volko, S. & Dong, X. The *Arabidopsis NPR1* gene that controls systemic acquired resistance encodes a novel protein containing ankyrin repeats. *Cell* **88,** 57–63 (1997).
8. Spoel, S. H. & Dong, X. How do plants achieve immunity? Defence without specialized immune cells. *Nature Rev. Immunol.* **12,** 89–100 (2012).
9. Pintard, L., Willems, A. & Peter, M. Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family. *EMBO J.* **23,** 1681–1687 (2004).

10. Spoel, S. H. *et al.* Proteasome-mediated turnover of the transcription coactivator NPR1 plays dual roles in regulating plant immunity. *Cell* **137,** 860–872 (2009).
11. Zhang, Y. *et al.* Negative regulation of defense responses in *Arabidopsis* by two *NPR1* paralogs. *Plant J.* **48,** 647–656 (2006).
12. Santner, A. & Estelle, M. Recent advances and emerging trends in plant hormone signalling. *Nature* **459,** 1071–1078 (2009).
13. Tan, X. *et al.* Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446,** 640–645 (2007).
14. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1-JAZ co-receptor. *Nature* **468,** 400–405 (2010).
15. Métraux, J.-P. *et al.* in *Advances in Molecular Genetics of Plant-Microbe Interactions* Vol. 1 (eds Hennecke, H. & Verma, D. P. S.) 432–439 (Kluwer Academic Publishers, 1991).
16. Mou, Z., Fan, W. & Dong, X. Inducers of plant systemic acquired resistance regulate NPR1 function through redox changes. *Cell* **113,** 935–944 (2003).
17. Nawrath, C., Heck, S., Parinthawong, N. & Métraux, J.-P. EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in *Arabidopsis*, is a member of the MATE transporter family. *Plant Cell* **14,** 275–286 (2002).
18. Wildermuth, M. C., Dewdney, J., Wu, G. & Ausubel, F. M. Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature* **414,** 562–565 (2001).
19. Gaffney, T. *et al.* Requirement of salicylic acid for the induction of systemic acquired resistance. *Science* **261,** 754–756 (1993).
20. Malamy, J., Carr, J. P., Klessig, D. F. & Raskin, I. Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection. *Science* **250,** 1002–1004 (1990).
21. Enyedi, A. J., Yalpani, N., Silverman, P. & Raskin, I. Localization, conjugation, and function of salicylic acid in tobacco during the hypersensitive reaction to tobacco mosaic virus. *Proc. Natl Acad. Sci. USA* **89,** 2480–2484 (1992).
22. Dorey, S. *et al.* Spatial and temporal induction of cell death, defense genes, and accumulation of salicylic acid in tobacco leaves reacting hypersensitively to a fungal glycoprotein. *Mol. Plant Microbe Interact.* **10,** 646–655 (1997).
23. Torres, M. A., Jones, J. D. G. & Dangl, J. L. Pathogen-induced, NADPH oxidase-derived reactive oxygen intermediates suppress spread of cell death in *Arabidopsis thaliana*. *Nature Genet.* **37,** 1130–1134 (2005).
24. Lu, H. *et al.* Genetic analysis of *acd6-1* reveals complex defense networks and leads to identification of novel defense genes in *Arabidopsis. Plant J.* **58,** 401–412 (2009).
25. Rate, D. N. & Greenberg, J. T. The *Arabidopsis aberrant growth and death2 mutant* shows resistance to *Pseudomonas syringae* and reveals a role for NPR1 in suppressing hypersensitive cell death. *Plant J.* **27,** 203–211 (2001).
26. Dieterle, M. *et al.* Molecular and functional characterization of *Arabidopsis* Cullin 3A. *Plant J.* **41,** 386–399 (2005).
27. Mackey, D., Holt, B. F., Wiig, A. & Dangl, J. L. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis. Cell* **108,** 743–754 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Zhang for sharing the *npr3*, *npr4*, *npr3 npr4* and *npr1 npr3 npr4* mutants; J. Song for providing the NPR3 and NPR4 Y2H constructs; Z. Mou for providing the data on the NPR1–GFP protein levels in the *nahG* transgenic plants, P. Zhou for discussion of the work and for critiquing the manuscript. This work was supported by the Hargitt Fellowship (to Z.Q.F.), grants GM069594-05 (to X.D.), CA107134 (to N.Z.), T32GM008268-23 (to J.R.), Grants-in-Aid for Scientific Research (no. 23120520) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to Y.T.). and the Royal Society Uf090321 (to S.H.S.). N.Z. is a Howard Hughes Medical Institute investigator and X.D. is a Howard Hughes Medical Institute-Gordon and Betty Moore Foundation investigator.

Author Contributions Z.Q.F., S.Y., A.S., R.M. and S.H.S. conceived and discovered that NPR3 and NPR4 mediate NPR1 degradation. Z.Q.F., A.S. and S.Y. found that SA regulates the interactions between the NPR proteins. S.Y., W.W. and A.S. found that NPR3 and NPR4 can bind SA with different affinities. J.R. and N.Z. showed that purified NPR4 recombinant protein exists as a tetramer, which is competent in SA binding. Z.Q.F., W.W. and R.M. demonstrated that the *npr3* and *npr4* single and double mutants are impaired in ETI and SAR. N.O. and Y.T. observed *in situ* accumulation of NPR1(C82A)–GFP in response to *Psm* ES4326/*avrRpt2*. S.H.S. provided data on the detection of NPR1–GFP protein in *eds5* and *ics1* plants. X.D. designed the research and, together with Z.F., S.Y., W.W., S.H.S., A.S. and R.M., wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to X.D. (xdong@duke.edu).

METHODS

***Arabidopsis thaliana* mutants and transgenic lines.** *Arabidopsis thaliana* mutants (in ecotype Col-0) *npr3-1*, *npr4-3*, *npr3-1 npr4-3*, *npr1-1 npr3-1 npr4-3*, *npr3-2* (SALK_043055), *npr4-2* (SALK_098460) and *npr3-2 npr4-2* were provided by Y. Zhang[11]. *35S:NPR1-GFP* was introduced into the *npr1-2 npr3-1 npr4-3* background by crossing the *35S:NPR1-GFP* transgenic plants (in *npr1-2*) with the *npr3-1 npr4-3* plants. Homozygous plants were selected by genotyping.

Co-immunoprecipitation assay. Three-week-old plant sample was collected and ground in liquid N_2. Protein was extracted in the extraction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 0.2% Nonidet P-40, and inhibitors: 50 μg ml^{-1} -*N*-tosyl-L-phenylalanyl chloromethyl ketone (TPCK), 50 μg ml^{-1} -*N*-tosyl-L-leucine chloromethyl ketone (TLCK), 0.6 mM phenylmethylsulphonyl fluoride (PMSF) and 40 μM MG115). The extracts were then pre-cleared with 50 μl of Dynabeads Protein G (Invitrogen). After 1 μl of anti-GFP antibody (Abcam) was added to the extracts and incubated for 2 h, 50 μl of magnetic Dynabeads was added to the samples and incubated for another hour with gentle rocking. The magnetic Dynabeads were then washed three times using the protein extraction buffer, and bound proteins were eluted by heating the magnetic beads in the SDS sample buffer containing 100 mM DTT at 95 °C for 10 min. The NPR1–GFP and CUL3 proteins were detected by western blotting using an anti-GFP antibody (Clontech) and an anti-CUL3A antibody[26], respectively[10].

Pathogen infection. To test for the hypersensitive response, the avirulent pathogen *Psm* ES4326 carrying *avrRpt2* (attenuance $(D)_{600\,nm} = 0.02$) or *avrRpm1* ($D_{600\,nm} = 0.1$) and *Pst* carrying *avrRps4* ($D_{600\,nm} = 0.1$) or *avrRpt2* ($D_{600\,nm} = 0.02$) were infiltrated into 3–4-week-old leaves. Cell death was recorded 2–3 days after the infiltration. Ion leakage was recorded over time as described[27]. To test for SAR, two lower leaves of 3-week-old plants were pressure-infiltrated with 10 mM $MgCl_2$ or avirulent bacterial pathogen *Psm* ES4326 carrying *avrRpt2* ($D_{600\,nm} = 0.02$). Three days later, virulent bacterial pathogen *Psm* ES4326 ($D_{600\,nm} = 0.001$) was infiltrated into two upper leaves. Disease symptoms were monitored and bacterial growth was analysed 3 days after the inoculation[10].

Imaging of NPR1–GFP in infection site. The *35S:NPR1(C82A)-GFP* plants were inoculated with *Psm* ES4326 carrying *avrRpt2* ($D_{600\,nm} = 0.02$) and incubated for 11 h. Leaf tissues were mounted in 10% glycerol and viewed with a BIOREVO (Keyence) BZ-9000 fluorescence microscope. The GFP signal is monitored with an excitation wavelength of 472.5 nm and a bandpass 502.5 to 537.5 nm emission filter. Red chlorophylls were viewed with an excitation wavelength of 540 nm and a bandpass 573 to 613 nm emission filter. To obtain wide-field view (2 × 5 pictures), image stitching was performed by BZ-II Image Analysis Application. Experiments were repeated eight times.

Molecular cloning of NPRs. The coding regions of *NPR1*, *2*, *3* and *4* were amplified with PrimeSTAR HS DNA polymerase (Takara) using specific primers containing Gateway attB sites (Supplementary Table 1), and then cloned into the pDONR207 entry vector using the BP Clonase (Invitrogen) to create the NPR entry clones. After verification by sequencing, each of the clones was mobilized using the LR Clonase (Invitrogen) into the Gateway destination vectors pDEST-GBKT7 and pDEST-GADT7 for yeast transformation[28], the protein expression vectors pDEST15 (Invitrogen) and pDEST-HisMBP (Addgene plasmid 11085)[29] for making the GST and His–MBP fusions, respectively.

Detection of the NPR1 protein. Four-week-old plants were sprayed with 0.5 mM SA and collected at different time points. Total protein was extracted in a buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.1% Triton X-100, 0.2% Nonidet P-40, and inhibitors: 50 μg ml^{-1} TPCK, 50 μg ml^{-1} TLCK, 0.6 mM PMSF and 40 μM MG115 (ref. 30). The homogenates were centrifuged twice at 13,500*g* for 15 min each. Protein was denatured in the SDS sample buffer containing 100 mM DTT at 75 °C for 10 min, and western blot analysis was performed using an antibody against NPR1 (ref. 16).

Quantitative real-time PCR. Total RNA was extracted from 4-week-old control and SA-treated plants at the indicated time points using TRIzol Reagent (Invitrogen). Genomic DNA was eliminated by treatment of the RNA with 2 U of TURBO DNA-free (Ambion). cDNA was synthesized using the Superscript III Reverse Transcription kit (Invitrogen) and analysed by quantitative real-time PCR using the FastStart Universal SYBR Green Master (Rox) kit (Roche) with gene-specific primers for *NPR1* and *ubiquitin 5* (Supplementary Table 1).

***In vitro* degradation assay.** The NPR1 degradation assay was performed as described[10]. Leaves from wild-type or *npr3-1 npr4-3* double mutant plants were ground in liquid N_2 and resuspended in the proteolysis buffer containing 25 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM NaCl, 10 mM ATP and 5 mM DTT. After centrifugation, the supernatants were mixed with the GST–NPR1 protein purified from *Escherichia coli* and incubated at room temperature for 15 min. The reactions were stopped by adding the SDS sample buffer containing 100 mM DTT and incubated at 75 °C for 10 min. The level of GST–NPR1 protein was analysed by western blotting using an anti-GST antibody (GE Healthcare).

***In vitro* pull-down assay.** The coding sequence of *CUL3A* was cloned into the GST vector pGEX4T-2 (GE Healthcare) for expression in *E. coli* BL21(DE3). The coding sequences of *NPR3* and *NPR4* were cloned into pCMX-PL2 for *in vitro* translation using TNT Quick Coupled Transcription/Translation System (Promega) to produce the HA-tagged NPR3 and NPR4 proteins. The purified GST–CUL3A protein was bound to the glutathione agarose beads, incubated with the HA-tagged NPR3 or NPR4 protein, and washed three times with EB buffer (50 mM Tris-HCl, pH 7.2, 100 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% dimethylsulphoxide, 20 mM DTT and 0.1% NP40). The HA-tagged NPR3 or NPR4 protein bound to GST–CUL3A protein was detected by western blot analysis using an anti-HA antibody (GenScript). Recombinant His-MBP-tagged NPR1 and GST-tagged NPR3 and NPR4 proteins were produced in *E. coli*. The recombinant His-MBP-tagged NPR1 was purified using the Ni-NTA resin (QIAGEN). GST-tagged NPR3 and NPR4 proteins were purified using glutathione beads and retained on the beads to pull down purified His–MBP–NPR1 protein with or without 100 μM SA in a buffer containing 50 mM Tris-HCl, pH 6.8, 100 mM NaCl and 0.1% NP40. After washing three times, bound His–MBP–NPR1 was eluted by heating the glutathione beads at 95 °C for 10 min in the SDS buffer with 100 mM DTT and detected by western blot analysis using an anti-His antibody (GenScript). Equal loadings were verified by staining the membrane with Ponceau S.

Yeast two-hybrid assay. The *Saccharomyces cerevisiae* yeast strains AH109 and Y187 were transformed with pGADT7-NPR1, 2, 3, 4 and pGBKT7-NPR1, 2, 3, 4, respectively, according to the Clontech yeast transformation protocol. Yeast strains were grown on SD−Trp−Leu plates, and then fresh single colonies were grown for 1 day in the SD−Trp−Leu liquid media. Interactions between bait and prey were detected on the selective media: SD−Trp−Leu−His (control), SD−Trp−Leu−His with 100 μM sodium salicylate (SA), 100 μM INA, or 100 μM 4-HBA. All of the SD−Trp−Leu−His selective media also contained 3 mM 3-aminotriazole.

SA-binding assay. The SA-binding assay was performed as described[6] with modifications. The GST–NPR3 and GST–NPR4 proteins were expressed in *E. coli* C41 and purified using Pierce Glutathione Magnetic Beads (Thermo). The protein-bound beads were incubated in 100 μl buffer containing 30 mM sodium citrate, pH 6.3, 1 mM EDTA and [^{3}H]-SA (American Radiolabelled Chemicals, specific activity 30 Ci mmol^{-1}). The beads were washed twice, resuspended in 100 μl H_2O, mixed with 6 ml Ultima Gold Cocktails (PerkinElmer) and counted using the LC6000SC liquid scintillation counter (Beckman Instruments). The non-specific binding was determined in the presence of 1 mM unlabelled SA. The data were analysed using GraphPad Prism 5.

Gel filtration analyses of NPR4. NPR4 was overexpressed as a GST-fusion protein in insect cells and purified by glutathione affinity chromatography in the presence of 100 mM DTT. After TEV cleavage, NPR4 was further purified by anion exchange chromatography and dialysed against a buffer containing 20 mM Tris-HCl, pH 8.0, 200 mM NaCl and 5 mM DTT. After concentration, 0.5–1 mg NPR4 was analysed in the same buffer with and without 10 μM SA as indicated by size exclusion chromatography on a Supderdex 200 gel filtration column. Co-elution of SA and NPR4 was monitored by [^{3}H]-SA, which was pre-mixed with the NPR4 protein before injection.

28. Rossignol, P., Collier, S., Bush, M., Shaw, P. & Doonan, J. H. *Arabidopsis* POT1A interacts with TERT-V(I8), an N-terminal splicing variant of telomerase. *J. Cell Sci.* **120,** 3678–3687 (2007).
29. Nallamsetty, S., Austin, B. P., Penrose, K. J. & Waugh, D. S. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* **14,** 2964–2971 (2005).
30. Fan, W. & Dong, X. *In vivo* interaction between NPR1 and transcription factor TGA2 leads to salicylic acid-mediated gene activation in *Arabidopsis*. *Plant Cell* **14,** 1377–1389 (2002).

LETTER

doi:10.1038/nature11073

The intense starburst HDF 850.1 in a galaxy overdensity at $z \approx 5.2$ in the Hubble Deep Field

Fabian Walter[1,2], Roberto Decarli[1], Chris Carilli[2,3], Frank Bertoldi[4], Pierre Cox[5], Elisabete Da Cunha[1], Emanuele Daddi[6], Mark Dickinson[7], Dennis Downes[5], David Elbaz[6], Richard Ellis[8], Jacqueline Hodge[1], Roberto Neri[5], Dominik A. Riechers[8], Axel Weiss[9], Eric Bell[10], Helmut Dannerbauer[11], Melanie Krips[5], Mark Krumholz[12], Lindley Lentati[3], Roberto Maiolino[3,13], Karl Menten[9], Hans-Walter Rix[1], Brant Robertson[14], Hyron Spinrad[15], Dan P. Stark[14] & Daniel Stern[16]

The Hubble Deep Field provides one of the deepest multiwavelength views of the distant Universe and has led to the detection of thousands of galaxies seen throughout cosmic time[1]. An early map of the Hubble Deep Field at a wavelength of 850 micrometres, which is sensitive to dust emission powered by star formation, revealed the brightest source in the field, dubbed HDF 850.1 (ref. 2). For more than a decade, and despite significant efforts, no counterpart was found at shorter wavelengths, and it was not possible to determine its redshift, size or mass[3–7]. Here we report a redshift of $z = 5.183$ for HDF 850.1, from a millimetre-wave molecular line scan. This places HDF 850.1 in a galaxy overdensity at $z \approx 5.2$, corresponding to a cosmic age of only 1.1 billion years after the Big Bang. This redshift is significantly higher than earlier estimates[3,4,6,8] and higher than those of most of the hundreds of submillimetre-bright galaxies identified so far. The source has a star-formation rate of 850 solar masses per year and is spatially resolved on scales of 5 kiloparsecs, with an implied dynamical mass of about 1.3×10^{11} solar masses, a significant fraction of which is present in the form of molecular gas. Despite our accurate determination of redshift and position, a counterpart emitting starlight remains elusive.

We have obtained a full-frequency scan of the 3-mm band towards the Hubble Deep Field using the IRAM (Institut de Radioastronomie Millimétrique) Plateau de Bure Interferometer. The observations covered the frequency range from 80–115 GHz in ten frequency settings at uniform sensitivity and at a resolution (about 2.3′′) that is a good match to galaxy sizes at high redshift. They resulted in the detection of two lines of carbon monoxide (CO), the most common tracer for molecular gas at high redshift[9], at 93.20 GHz and 111.84 GHz at the position of HDF 850.1. Identifying these lines with the $J = 5$ and $J = 6$ rotational transitions of CO gives a redshift for HDF 850.1 of $z = 5.183$. This redshift was then unambiguously confirmed by the Plateau de Bure Interferometer's detection of the 158-μm line of ionized carbon ([C II], redshifted to 307.38 GHz), one of the main cooling lines of the star-forming interstellar medium. Stacking of other molecules covered by our frequency scan that trace higher volume densities did not lead to a detection (see Supplementary Information). Subsequently, the $J = 2$ line of CO has also been detected using the National Radio Astronomy Observatory (NRAO) Jansky Very Large Array at 37.29 GHz. The observed [C II] and CO spectra towards HDF 850.1 are shown in Fig. 1.

The beam size of our CO observations (about 2.3′′, 15 kpc at $z = 5.183$) is too large to spatially resolve the molecular gas emission in HDF 850.1. However, the [C II] and underlying continuum observations (around 1.2′′ × 0.8′′) show that the source is extended (hitherto, the interstellar medium has been spatially resolved only in

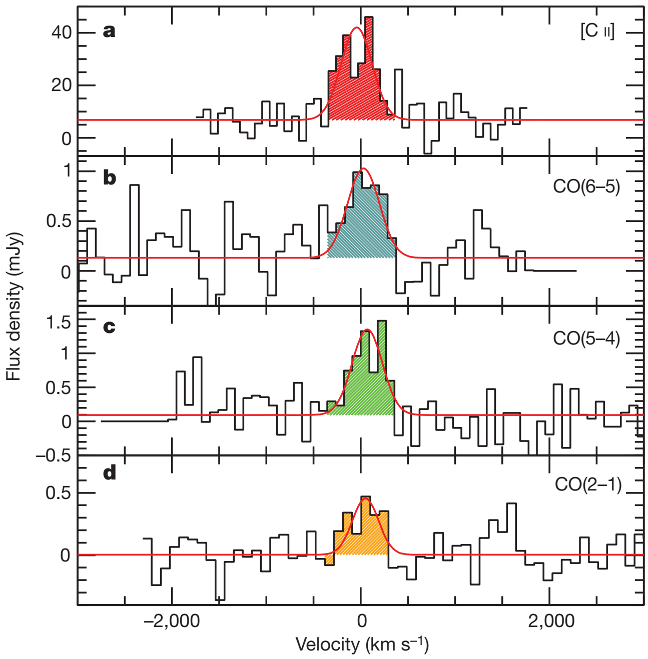


Figure 1 | Detection of four lines tracing the star-forming interstellar medium in HDF 850.1. **a**, [C II], $\nu_{obs} = 307.383$ GHz. **b**, CO(6–5), $\nu_{obs} = 111.835$ GHz. **c**, CO(5–4), $\nu_{obs} = 93.202$ GHz. **d**, CO(2–1), $\nu_{obs} = 37.286$ GHz. Zero velocity corresponds to a redshift of z = 5.183. Continuum emission is detected in **a** and **b** at 6.80 ± 0.8 mJy and 0.13 ± 0.03 mJy, respectively. We derive a 3σ continuum limit of 30 μJy from the Jansky Very Large Array observations at 37.3 GHz using a bandwidth larger than shown here. Gaussian fits to the lines give a full width at half maximum (FWHM) of 400 ± 30 km s^{-1}, narrower than typically found in sub-millimetre selected galaxies[13]. The observed integrated line flux densities are: $S_{[C\,II]} = 14.6 \pm 0.3$ Jy km s^{-1}, $S_{[CO(6-5)]} = 0.39 \pm 0.1$ Jy km s^{-1}, $S_{[CO(5-4)]} = 0.50 \pm 0.1$ Jy km s^{-1} and $S_{[CO(2-1)]} = 0.17 \pm 0.04$ Jy km s^{-1}. The resulting line luminosities are[9] 5.0×10^{10} K km s^{-1} pc^2, 1.0×10^{10} K km s^{-1} pc^2, 1.9×10^{10} K km s^{-1} pc^{2}and 4.1×10^{10} K km s^{-1} pc^2 or $1.10 \times 10^{10} L_{Sun}$, $1.06 \times 10^8 L_{Sun}$, $1.14 \times 10^8 L_{Sun}$ and $1.5 \times 10^7 L_{Sun}$ (uncertainties as given for integrated line flux densities). Large velocity gradient modelling gives a predicted CO(1–0) line luminosity of 4.3×10^{10} K km s^{-1} pc^2.

[1]Max-Planck Institut für Astronomie, Königstuhl 17, D-69117, Heidelberg, Germany. [2]National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, PO Box O, Socorro, New Mexico 87801, USA. [3]Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge CB3 OHE, UK. [4]Argelander Institute for Astronomy, University of Bonn, Auf dem Hügel 71, 53121 Bonn, Germany. [5]IRAM, 300 rue de la Piscine, F-38406 Saint-Martin d'Hères, France. [6]Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, Irfu/Service d'Astrophysique, CEA Saclay, Orme des Merisiers, 91191 Gif-sur-Yvette cedex, France. [7]National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, Arizona 85719, USA. [8]Astronomy Department, California Institute of Technology, MC105-24, Pasadena, California 91125, USA. [9]Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany. [10]Department of Astronomy, University of Michigan, 500 Church Street, Ann Arbor, Michigan 48109, USA. [11]Universität Wien, Institut für Astronomie, Türkenschanzstraße 17, 1080 Wien, Austria. [12]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. [13]INAF-Osservatorio Astronomico di Roma, via di Frascati 33, 00040 Monte Porzio Catone, Italy. [14]Department of Astronomy, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. [15]Department of Astronomy, University of California at Berkeley, Berkeley, California 94720, USA. [16]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA.

extremely rare quasar host galaxies at such high redshifts[10]). A single Gaussian fit yields a deconvolved size of $0.9 \pm 0.3''$, or 5.7 ± 1.9 kpc, at the redshift of the source. Figure 2 shows the maps of total [C II] emission (Fig. 2a) as well as the red- and blue-shifted parts of the [C II] line (Fig. 2b) superposed on the deepest available Hubble Space Telescope images of the Hubble Deep Field[1]. The derived dynamical mass is $M_{dyn} \approx 1.3 \pm 0.4 \times 10^{11} M_{Sun}$, assuming an arbitrary inclination of 30°. An alternative interpretation is that the source is a merger of two galaxies, rather than a single rotating disk, which would lower the implied dynamical mass. Figure 2 shows that the source is completely obscured in the observed optical and near-infrared wavebands (that is, the rest-frame ultraviolet). There is no indication of HDF 850.1 harbouring an active galactic nucleus powered by a supermassive black hole (quasar)[11].

The CO(6–5)/CO(2–1) line luminosity ratio (in units of $\mathrm{K\,km\,s^{-1}\,pc^2}$) (ref. 9) is 0.23 ± 0.05. Assuming that the gas is being emitted from the same volume, this implies that the high-J CO emission is sub-thermally excited on galactic scales, less than seen in the nuclei of local starburst galaxies[12]. Using a standard large velocity gradient model we find that the observed CO line intensities can be fitted with a moderate molecular hydrogen density of $10^{3.2}\,\mathrm{cm^{-3}}$ and a kinetic temperature of 45 K for virialized clouds (velocity gradient $dv/dr = 1.2\,\mathrm{km\,s^{-1}\,pc^{-1}}$). We caution that these numbers would change if the CO transitions were not emitted from the same volume. The predicted CO(1–0) line luminosity is $4.3 \times 10^{10}\,\mathrm{K\,km\,s^{-1}\,pc^2}$, close to the measured value for CO(2-1). Depending on the choice of α, the CO-to-H_2 conversion factor, this line luminosity implies a molecular gas mass of $M_{H2} = 3.5 \times (\alpha/0.8) \times 10^{10} M_{Sun}$; here $\alpha = 0.8$, in units of $M_{sun}\,(\mathrm{K\,km\,s^{-1}\,pc^2})^{-1}$, is the conversion factor adopted for ultra-luminous infrared galaxies (ULIRGs)[13] and thought to be applicable to sub-millimetre bright objects[14]. The implied molecular gas mass fraction is $M_{H2}/M_{dyn} \sim 0.25 \pm 0.08\,(\alpha/0.8)$; that is, even with a low ULIRG conversion factor the molecular gas constitutes a significant fraction of the overall dynamical mass. This molecular gas mass (and fraction) is comparable to what is found in other sub-millimetre bright galaxies that are typically located at much lower redshift[14,15].

The line-free channels of the observations (Fig. 1) were used to constrain the underlying continuum emission. Our accurate position of the rest-frame 158 μm emission is indicated as a cross in Fig. 2 (right). We combine our continuum detections at 307 GHz and 112 GHz with published values and new Herschel Space Telescope observations to constrain the far-infrared properties of the source (see Supplementary Information for details). Our best fit gives a far-infrared luminosity of $L_{FIR} = (6.5 \pm 1) \times 10^{12} L_{Sun}$, a dust temperature of 35 ± 5 K (that is, broadly consistent with the average kinetic temperature of the molecular gas), a dust mass of $M_{dust} = (2.75 \pm 0.5) \times 10^8 M_{Sun}$ and a star formation rate of $850 M_{Sun}$ per year (with an uncertainty of about 30%). Given the extent of the source this results in an galaxy-averaged star formation rate surface density of $850 M_{Sun}$ per year divided by $(\pi(2.8\,\mathrm{kpc})^2)$ equalling $35 M_{Sun}$ per year $\mathrm{kpc^{-2}}$ (uncertainty ~50%), more than an order of magnitude less than found in nearby merging systems and a compact quasar host galaxy at $z = 6.42$ that has been studied in similar detail[10]. HDF 850.1 falls on the universal local star-formation law that relates the average surface density of the star formation rate to that of the molecular gas mass per local free-fall time[16]. The estimated surface density would increase if future observations resolved the source structure.

The resulting [C II]/far-infrared luminosity ratio of $L_{[C II]}/L_{FIR} = (1.7 \pm 0.5) \times 10^{-3}$ in HDF 850.1 is comparable to what is found in normal local star-forming galaxies[17], but is an order of magnitude higher than what is found in a $z = 6.42$ quasar[10], the only other high-z system where the [C II] emission could be resolved to date. Recent studies indicate that this ratio is a function of environment, with a low value ($L_{[C II]}/L_{FIR} \approx 1 \times 10^{-4}$) for luminous systems dominated by a central black hole (quasars) and a high ratio (up to $L_{[C II]}/L_{FIR} \approx 1 \times 10^{-2}$) for low-metallicity environments. Our relatively high ratio in $L_{[C II]}/L_{FIR}$ is consistent with HDF 850.1 being a high redshift star-forming system in a non-quasar environment[17].

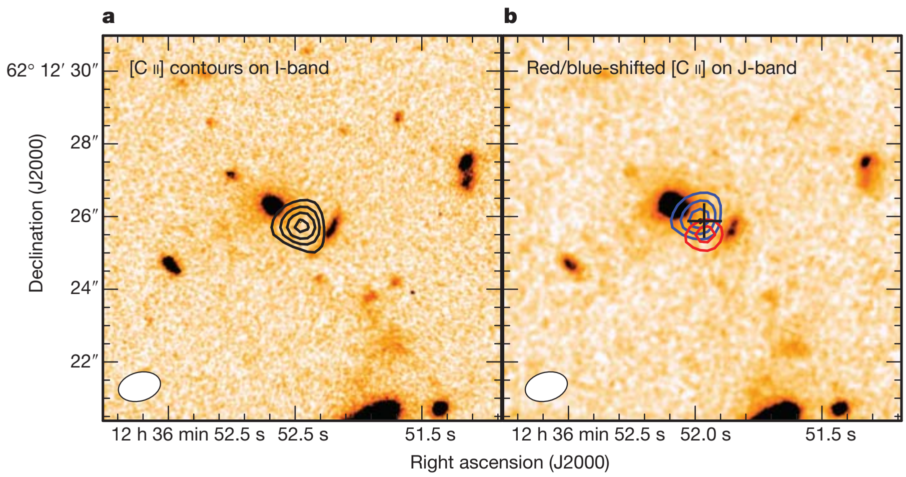


Figure 2 | [C II] line emission towards HDF 850.1. a, [C II] contours on top of a deep Hubble Space Telescope image[1] of the region in a filter (I band) that covers the Lyman-α line and ultraviolet continuum at $z = 5.183$. [C II] contours show the averaged emission over $700\,\mathrm{km\,s^{-1}}$ and are plotted at 5 mJy per beam, 7 mJy per beam, 9 mJy per beam and 11 mJy per beam ($1\sigma = 1.3$ mJy per beam). A Gaussian fit to the emission gives a deconvolved size of $0.9 \pm 0.3''$ or 5.7 ± 1.9 kpc at $z = 5.183$. The underlying continuum emission (not shown) is also extended on the same scales. **b**, The blue and red contours indicate the approaching and receding [C II] emission relative to the systemic redshift of $z = 5.183$. The colour shows a deep Hubble Space Telescope image in a longer wavelength filter (the J band from the Hubble Space Telescope's near-infrared camera and multi-object spectrometer (NICMOS))[29]. The cross indicates the position (and its 5σ uncertainty) of the rest-frame 158-μm continuum emission peak (right ascension 12 h 36 min 51.976 s, declination 62° 12′ 25.80″ in the J2000.0 system), consistent with earlier millimetre interferometric measurements[3,6] at lower resolution. The [C II] contours have been derived by averaging the spectrum (Fig. 1) from $-400\,\mathrm{km\,s^{-1}}$ to $0\,\mathrm{km\,s^{-1}}$ and $0\,\mathrm{km\,s^{-1}}$ to $+400\,\mathrm{km\,s^{-1}}$ and are plotted at levels of 7 mJy per beam, 10 mJy per beam and 13 mJy per beam ($1\sigma = 1.8$ mJy per beam), respectively. In each panel the beam size of the [C II] observations ($1.23'' \times 0.81''$) is indicated in the bottom left corner. From the spatial offset (total offset = 0.9″, that is, radius r is 0.45″ or 2.8 kpc) and the FWHM of the line, we derive an approximate dynamical mass of $M_{dyn} \approx 3.4 \times 10^{10} M_{Sun}/(\sin i)^2$ where i is the (unknown) inclination of the system (using $M_{dyn} \sin^2 i = 1.3 \times (\mathrm{FWHM}/2)^2 r/G$, where G is the gravitational constant[30]). These deep Hubble Space Telescope images of the Hubble Deep Field fail to reveal the (rest-frame) ultraviolet/optical counterpart of the galaxy that is forming stars at a rate of about $850 M_{Sun}$ per year.

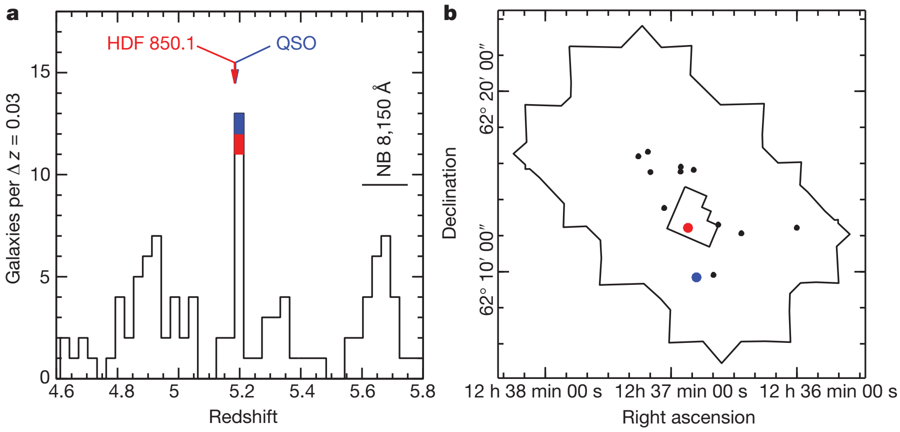


Figure 3 | Distribution of galaxies near HDF 850.1. **a**, Distribution of spectroscopic redshifts towards the Hubble Deep Field and its surroundings (from the Great Observatories Origins Deep Survey-North, GOODS-N). HDF 850.1 is indicated in red, and the quasar at the same redshift[18] is indicated in blue. There is an overdensity of galaxies in the redshift bin that contains HDF 850.1. The high source density at $z \approx 5.7$ is an observational artefact due to narrow-band Lyman-α imaging surveys of the region (with spectroscopic follow-up) that are sensitive to this particular narrow redshift range. **b**, Spatial coverage of the sources in the redshift bin $z = 5.183$–5.213. The small border indicates the size of the Hubble Deep Field; the larger border shows the surrounding area of GOODS-N. The presence of a strongly star-forming galaxy (HDF 850.1) and a quasar[18] in this region provides evidence for cosmic structure formation in the first billion years of the Universe. See Supplementary Information for more details.

An inspection of the distribution of galaxies towards HDF 850.1 that have spectroscopic redshifts shows that there is an overdensity of galaxies at the exact redshift of HDF 850.1, including a quasar at $z = 5.186$[18] (Fig. 3 and Supplementary Information). This makes this region one of the most distant galaxy overdensities known to date[19]. An elliptical galaxy at $z = 1.224$ (ref. 20) that is situated close to HDF 850.1 in projection (around 1′′ to the northeast) could potentially act as a gravitational lens for this source[3,4,21]. Using a velocity dispersion of 146 km s^{-1} in a singular isothermal sphere for this elliptical galaxy[4] and our new redshift and position of HDF 850.1, we derive an amplification factor of around 1.4. A similar flux amplification is found for a simple point source lens model with mass $3.5 \times 10^{11} M_{\mathrm{Sun}}$. This implies that even if lensing is occurring, the quantities derived here would not need to be revised significantly.

HDF 850.1 remains outstanding in the study of dust-obscured starbursts at high redshift, being one of the first such sources discovered, and yet evading detection in the optical and near-infrared. Its redshift of $z = 5.183$ enforces the presence of a high redshift tail ($z > 4$) of submillimetre bright star-forming galaxies (that is, a galaxy without an active galactic nucleus); currently there are only about half a dozen systems known[22–26]. Only a small fraction of submillimetre-bright sources is expected to be at very high redshift[27]—it is thus ironic that the first blank-field source belongs to this subgroup. HDF 850.1's large spatial extent, in combination with the modest CO excitation, the moderate surface density of its star-formation rate, and a high [C II]/far-infrared luminosity ratio, points to the presence of a spatially extended major starburst that is completely obscured even in the deepest Hubble Space Telescope images available for the Hubble Deep Field. The absence of a possible counterpart in the available deep imaging, even though the star-forming interstellar medium is distributed over many square kiloparsecs, makes this source extreme[22–24]. Given its high molecular gas mass ($3.5 \times (\alpha/0.8) \times 10^{10} M_{\mathrm{Sun}}$) and star-formation rate ($850 M_{\mathrm{Sun}}$ per year), HDF 850.1 can build a significant stellar component as early as $z \approx 4$ (ref. 28; a few hundred million years from $z \approx 5$). Blind line searches through spectral scans at millimetre wavelengths, as performed here, thus play a fundamental role in unveiling the nature of star-forming galaxies that are completely obscured in the (restframe) optical and ultraviolet even if multiwavelength data at unparalleled depth are available.

Received 23 December 2011; accepted 22 March 2012.

1. Williams, R. E. *et al.* The Hubble Deep Field: observations, data reduction, and galaxy photometry. *Astron. J.* **112,** 1335–1389 (1996).
2. Hughes, D. H. *et al.* A submillimetre survey of the Hubble Deep Field: unveiling dust-enshrouded star formation in the Early Universe. *Nature* **394,** 241–247 (1998).
3. Downes, D. *et al.* Proposed identification of Hubble Deep Field submillimeter source HDF 850.1. *Astron. Astrophys.* **347,** 809–820 (1999).
4. Dunlop, J. S. *et al.* Discovery of the galaxy counterpart of HDF 850.1, the brightest submillimetre source in the Hubble Deep Field. *Mon. Not. R. Astron. Soc.* **350,** 769–784 (2004).
5. Wagg, J. *et al.* A broad-band spectroscopic search for CO line emission in HDF850.1: the brightest submillimetre object in the Hubble Deep Field-north. *Mon. Not. R. Astron. Soc.* **375,** 745–752 (2007).
6. Cowie, L. L., Barger, A. J., Wang, W.-H. & Williams, J. P. An accurate position for HDF 850.1: the brightest submillimeter source in the Hubble Deep Field-north. *Astrophys. J.* **697,** L122–L126 (2009).
7. Carilli, C. L. & Yun, M. S. The radio-to-submillimeter spectral index as a redshift indicator. *Astrophys. J.* **513,** L13–L16 (1999).
8. Richards, E. A. Radio Identification of Submillimeter Sources in the Hubble Deep Field. *Astrophys. J.* **513,** L9–L12 (1999).
9. Solomon, P. M. & Vanden Bout, P. A. Molecular Gas at High Redshift. *Annu. Rev. Astron. Astrophys.* **43,** 677–725 (2005).
10. Walter, F. *et al.* A kiloparsec-scale hyper-starburst in a quasar host less than 1 gigayear after the Big Bang. *Nature* **457,** 699–701 (2009).
11. Alexander, D. *et al.* The Chandra Deep Field North survey. XIII. 2 ms point-source catalogs. *Astron. J.* **126,** 539–574 (2003).
12. Loenen, A. F. *et al.* Excitation of the molecular gas in the nuclear region of M 82. *Astron. Astrophys.* **521,** L2 (2010).
13. Downes, D. & Solomon, P. M. Rotating nuclear rings and extreme starbursts in ultraluminous galaxies. *Astrophys. J.* **507,** 615–654 (1998).
14. Tacconi, L. *et al.* Submillimeter galaxies at z ~ 2: evidence for major mergers and constraints on lifetimes, IMF, and CO-H2 conversion factor. *Astrophys. J.* **680,** 246–262 (2008).
15. Ivison, R. *et al.* Tracing the molecular gas in distant submillimetre galaxies via CO(1–0) imaging with the Expanded Very Large Array. *Mon. Not. R. Astron. Soc.* **412,** 1913–1925 (2011).
16. Krumholz, M. R., Dekel, A. & McKee, C. F. A universal, local star formation law in galactic clouds, nearby galaxies, high-redshift disks, and starbursts. *Astrophys. J.* **745,** 69 (2012).
17. Stacey, G. J. *et al.* A 158 μm [C II] line survey of galaxies at z ~ 1–2: an indicator of star formation in the early Universe. *Astrophys. J.* **724,** 957–974 (2010).
18. Barger, A. J. *et al.* X-ray, optical, and infrared imaging and spectral properties of the 1Ms Chandra Deep Field North sources. *Astron. J.* **124,** 1839–1885 (2002).
19. Capak, P. *et al.* A massive protocluster of galaxies at a redshift of z ≈ 5.3. *Nature* **470,** 233–235 (2011).
20. Barger, A. J., Cowie, L. L. & Wang, W.-H. A highly complete spectroscopic survey of the GOODS-N field. *Astrophys. J.* **689,** 687–708 (2008).
21. Hogg, D. W., Blandford, R., Kundic, T., Fassnacht, C. D. & Malhotra, S. A candidate gravitational lens in the Hubble Deep Field. *Astrophys. J.* **467,** L73–L75 (1996).
22. Riechers, D. A. *et al.* A massive molecular gas reservoir in the z = 5.3 submillimeter galaxy AzTEC-3. *Astrophys. J.* **720,** L131–L136 (2010).
23. Daddi, E. *et al.* Two bright submillimeter galaxies in a z = 4.05 protocluster in Goods-North, and accurate radio-infrared photometric redshifts. *Astrophys. J.* **694,** 1517–1538 (2009).
24. Schinnerer, E. *et al.* Molecular gas in a submillimeter galaxy at z = 4.5: evidence for a major merger at 1 billion years after the Big Bang. *Astrophys. J.* **689,** L5–L8 (2008).
25. Combes, F. *et al.* A bright z = 5.2 lensed submillimeter galaxy in the field of Abell 773. HLSJ091828.6+514223. *Astron. Astrophys.* **538,** L4 (2012).

26. Cox, P. *et al.* Gas and dust in a submillimeter galaxy at z = 4.24 from the Herschel atlas. *Astrophys. J.* **740,** 63 (2011).
27. Ivison, R. *et al.* A robust sample of submillimetre galaxies: constraints on the prevalence of dusty, high-redshift starbursts. *Mon. Not. R. Astron. Soc.* **364,** 1025–1040 (2005).
28. Wiklind, T. *et al.* A population of massive and evolved galaxies at z>~5. *Astrophys. J.* **676,** 781–806 (2008).
29. Dickinson, M. *et al.* The unusual infrared object HDF-N J123656.3+621322. *Astrophys. J.* **531,** 624–634 (2000).
30. Daddi, E. *et al.* Very high gas fractions and extended gas reservoirs in z = 1.5 disk galaxies. *Astrophys. J.* **713,** 686–707 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is based on observations carried out with the IRAM Plateau de Bure Interferometer. IRAM is supported by MPG (Germany), INSU/CNRS (France) and IGN (Spain). The Jansky Very Large Array of NRAO is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. D.A.R. acknowledges support from NASA through a Spitzer Space Telescope grant. R.D. acknowledges funding through DLR project FKZ 50OR1004.

Author Contributions F.W. had the overall lead of the project. The Plateau de Bure Interferometer data were analysed by R.D., F.W., P.C., R.N., M.K. and D.D. The Jansky Very Large Array data reduction was performed by C.C., J.H. and L.L. The molecular gas excitation was led by A.W. Spectroscopic redshift information was provided by M.D., R.E., H.S., D.S. and D.P.S. The spectral energy distribution analysis, including new Herschel data, was led by E.D.C, D.E. and E.D. An updated lensing model was provided by D.D. All authors helped with the proposal, data analysis and interpretation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.W. (walter@mpia.de).

LETTER

doi:10.1038/nature11165

Possible tropical lakes on Titan from observations of dark terrain

Caitlin A. Griffith[1], Juan M. Lora[1], Jake Turner[1], Paulo F. Penteado[2], Robert H. Brown[1], Martin G. Tomasko[1], Lyn Doose[1] & Charles See[1]

Titan has clouds, rain and lakes—like Earth—but composed of methane rather than water. Unlike Earth, most of the condensable methane (the equivalent of 5 m depth globally averaged[1]) lies in the atmosphere. Liquid detected on the surface (about 2 m deep) has been found by radar images only poleward of 50° latitude[2,3], while dune fields pervade the tropics[4]. General circulation models explain this dichotomy, predicting that methane efficiently migrates to the poles from these lower latitudes[5–7]. Here we report an analysis of near-infrared spectral images[8] of the region between 20° N and 20° S latitude. The data reveal that the lowest fluxes in seven wavelength bands that probe Titan's surface occur in an oval region of about 60 × 40 km^2, which has been observed repeatedly since 2004. Radiative transfer analyses demonstrate that the resulting spectrum is consistent with a black surface, indicative of liquid methane on the surface. Enduring low-latitude lakes are best explained as supplied by subterranean sources (within the last 10,000 years), which may be responsible for Titan's methane, the continual photochemical depletion of which furnishes Titan's organic chemistry[9].

At narrow wavelength regions, or 'windows', centred at 0.93 μm, 1.08 μm, 1.28 μm, 1.58 μm, 2.00 μm, 2.80 μm and 5.0 μm, Titan's atmosphere is sufficiently transparent to allow visibility of the surface[9,10]. Within 20° S and 20° N latitude, the optical depth and the scattering properties of the two main sources of opacity, methane and photochemical haze, are well characterized, both by *in situ* measurements from the Huygens probe at 10° S latitude and 191° W longitude[11,12], and remote Cassini measurements[13,14]. The determined constraints enable radiative-transfer analyses of Titan's spectra to discern the effects of the surface from those of the atmosphere. We consider data obtained only at low phase angles ($<32°$), low incident and scattering angles ($<45°$), and a pixel scale less than 15 km per pixel, such that surface terrains are best discerned and most directly illuminated and viewed. Given these specifications and an exposure time of at least 50 s, between October 2004 and December 2008, Cassini's Visual and Infrared Mapping Spectrometer (VIMS) imaged about 17% (2.85×10^6 km^2) of Titan's 20° S–20° N surface, with the landing site most comprehensively sampled (Supplementary Fig. 1, Supplementary Information).

The selected data comprise 32,564 spectra, which display values for Titan's outgoing intensity divided by its incident flux (I/F) that span an order of magnitude in the clearest windows. The general trend of these values is well understood. At the shortest-wavelength windows, I/F values are higher (Fig. 1), because the atmospheric scattering of sunlight from the bright haze increases with decreasing wavelengths[11]. Yet the unique statistics of the dark terrains is unexpected: the lowest I/F values occur in the same pixels (same terrain) for all windows; that is, they represent the same observed spectrum. In addition, these pixels form an abrupt lower threshold to the range of I/F values, which is in contrast to the gradual falling off of the highest I/F values (Fig. 1). The sudden lower cut-off in the I/F and its wavelength dependence suggest that we have found the signature of a black surface, where the observed I/F is controlled solely by the scattering and absorption of sunlight due to the overlying atmosphere. The lowest I/F values describe a contiguous oval region at 14° S latitude and 173° W longitude, covering an area of 40×60 km^2 defined by 21 spectra, which are identical to within the noise (Figs 2 and 3). This 'dark oval' is observed on the TA, T10 and T35 Titan fly-bys of October 2004, January 2006, and August 2007. It is unusual, covering an area 0.05% of that sampled. Within 7% of this I/F are four slightly brighter terrains.

We derive the surface albedos of the dark oval feature using a discrete-ordinates radiative-transfer analysis with 32 streams (Fig. 3). This calculation was verified to correctly reproduce the surface albedo of the Huygens probe landing site, which was determined (at 0.4–1.6 μm) from *in situ* measurements by the Huygens Descent Imager/Spectral Radiometer (DISR). Radiative-transfer analyses of the dark oval spectra yield surface albedos consistent, at all windows, with a black surface (Fig. 3). The compounded uncertainties in the derived surface albedos, resulting from the VIMS calibration, the atmospheric

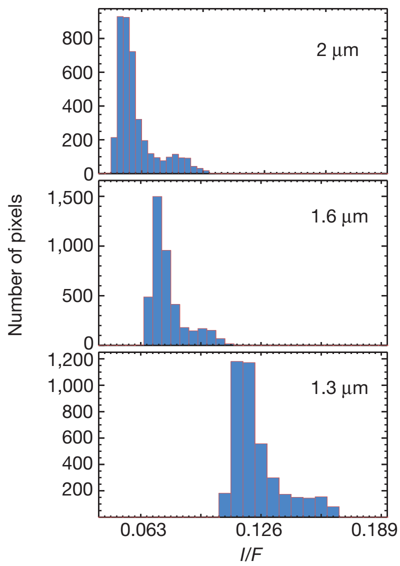


Figure 1 | Histograms of Titan's *I/F* values at three windows. Values derive from the 4,096 pixels of the VIMS V1567239055 spectral image. The lowest I/F values include the dark oval feature (Figs 2 and 3) and represent abrupt lower limits to Titan's I/F, consistent with that caused solely by scattering of the moon's haze. Such an abrupt cut-off, not typical of Titan's bright terrain (Supplementary Information), suggests a non-reflective surface, and calls for a radiative transfer analysis of the spectra. The detection of a black surface is particularly interesting, because Cassini radar images detect surface liquids through the presence of a black surface at 2.2 cm, which, at the 2% albedo level, requires a liquid depth of 8 m (refs 2, 17). Cassini near-infrared observations are more sensitive to the presence of shallow methane lakes than are Cassini radar observations, because methane liquid absorbs more strongly in the near-infrared window wavelengths, compared to at 2.2 cm. In principle, the variable cross-section of liquid methane in the near-infrared windows can constrain the depths of lakes from about 1 cm to about 2 m deep.

[1]Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA. [2]Universidade de São Paulo, IAG, Rua do Matao 1226, São Paulo, SP 05508-090, Brazil.

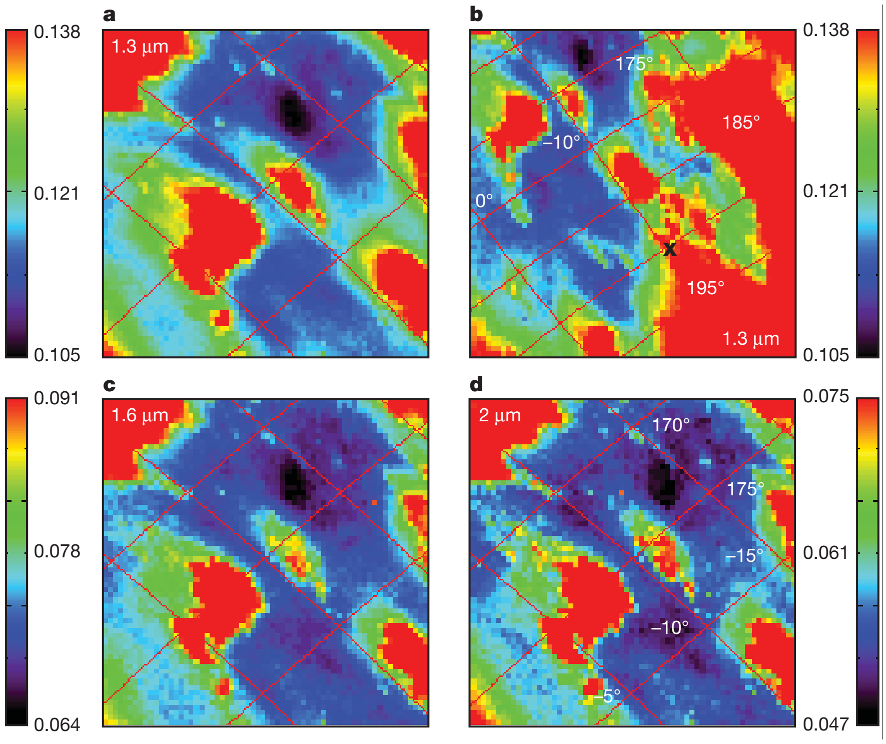


Figure 2 | Images of Titan at three wavelengths that probe the surface. **a**, **c** and **d**, *I*/*F* values at 1.3 µm, 1.6 µm and 2.0 µm from the VIMS V1567239055 spectral image of the T35 fly-by indicate a feature at $-14°$ latitude and 173° W longitude, defined by 21 spectra. This feature also appears in a lower-resolution image, V1567241480 (**b**), that includes the probe landing site (marked with an X). Data were reduced with the standard VIMS flats, which were checked for consistency with images of the haze and its north–south asymmetry. The *I*/*F* values in these and other windows are used to derive surface albedos, the uncertainty of which increases with decreasing wavelengths. This trend occurs because sensitivity of *I*/*F* to the surface albedo increases with wavelength, owing to the decreasing atmospheric scattering efficiency. Uncertainties in the surface albedos, caused mainly by the calibration of the VIMS instrument, the uncertainties in the atmospheric opacity, and the noise, are discussed in the Supplementary Information.

opacity, and noise, are 0.087, 0.036, 0.023, 0.020, 0.022, 0.021 and 0.011, at the window wavelengths of 0.93 µm, 1.08 µm, 1.28 µm, 1.58 µm, 2.00 µm, 2.78 µm and 5.0 µm, respectively. Errors at the shortest four wavelengths are predominantly the 1σ error in the VIMS calibration; those of the three longest wavelengths are due to uncertainties in the atmospheric opacity (Supplementary Information). The 0.93-µm window does not place strong constraints on the surface albedo, and therefore is discussed only in the consideration of relative surface albedos. These values agree with the uncertainties estimated independently from the comparison of the VIMS and DISR derivations of the landing site surface albedos[14] (Supplementary Information). Our finding of a surface with an albedo smaller at all windows than the above stated uncertainties is consistent with the 5-µm albedo of the south polar lake Ontario Lacus, also determined to be black[15], from constraints on the scattered light of the atmosphere determined through a comparison of *I*/*F* measurements made at different emission angles. The spectra indicate no evidence for an ethane feature, as detected in Ontario Lacus[15]: the ratio of spectra in and out of the dark oval resembles features caused by the ratio of spectra of any dark to bright terrain at the same lighting angles, and its width and shape are consistent with the transparency of the window. Therefore, our spectral ratios, while compositionally inconclusive, are consistent with a black surface at all windows, since spectroscopic features require backscattered light.

The absorption cross-section of liquid methane varies by several orders of magnitude, such that only 2 m is needed to cause a black surface albedo (to within a value of 0.02) in all windows, but less methane (for example, depths of 20 cm and 1 mm at wavelengths of 1.58 µm and 5.0 µm respectively) blackens particular windows[16]. Liquid methane deeper than about 1 m, depending on the albedo of the lake bottom, readily explains the dark oval's surface albedos. Such surface optical properties have not been detected on any other moon, and are thus inconsistent with any of the solid terrains thus represented. Fresnel reflection from a lake contributes to *I*/*F* at short wavelengths, where the surface is diffusely illuminated, but only by at most 2% (the value of an isotropically illuminated lake), which is within the error of our measurements. Particulates in liquid methane increase the lake-surface albedo. While too little is known to predict particle concentrations, a simple estimate of their effects suggests that a lake on Titan is black (to within 2%) for expected haze concentrations, assuming nearly spherical particles, and for sand-sized particles with concentrations smaller than $10^{-6}\,m^{-3}$, which is typical of terrestrial lakes. Mudflats and hyper-concentrated flows, such as those in terrestrial desert washes following a rainstorm, are needed to brighten the near-infrared albedo of methane lakes significantly (Supplementary Information).

Our survey identified four other terrains, two of which are defined by more than three pixels, that have *I*/*F* values at 1.58 µm, 2.0 µm, 2.8 µm and 5 µm similar to that of the dark oval, within the uncertainties (Supplementary Table 2). Yet their *I*/*F* values at 0.938 µm, 1.08 µm and 1.28 µm are, in contrast, significantly higher than those of the dark oval, considering the relative uncertainties in the surface albedos, which are estimated to be lower than the absolute albedo determinations (Supplementary Figs 2 and 3, and Supplementary Information). Given the absorptive properties of liquid methane, the spectra of these brighter terrains are consistent with a surface dampened by 1–9 cm of liquid, depending on the albedo of the underlying surface. Such a wet surface, more transparent at 2.2 cm, would be difficult to distinguish using radar because it attenuates the 2.2 cm reflectivity by less than 5% (ref. 17). One of these brighter regions (at 7° S, 185° W) is interesting because, unlike the dark oval region, it has been observed with radar measurements, which indicate that it lies in a dune field. Much of this dune field is 15% brighter than the dark oval region at all near-infrared wavelengths, and thus shows no evidence of liquid. Previous work finds that the *I*/*F* variability results from compositional changes in the broad inter-dune regions[18]. Whether inter-dune regions become wet, as occurs in the Namibian dunes on Earth, which share traits with Titan's dunes[4,19], can be tested with high-resolution images. That these low albedos occur in the dunes implies that wet terrain is associated with low-altitude regions. The detected five dark terrains (Supplementary Table 2) suggest that small lakes and swamps sparsely dot Titan's tropical surface; the determination of their depths and extents require additional near-infrared and radar measurements.

The dark oval region has been present since 2004, several years before the arrival of seasonal tropical clouds[20–22]. Its presence during the dry

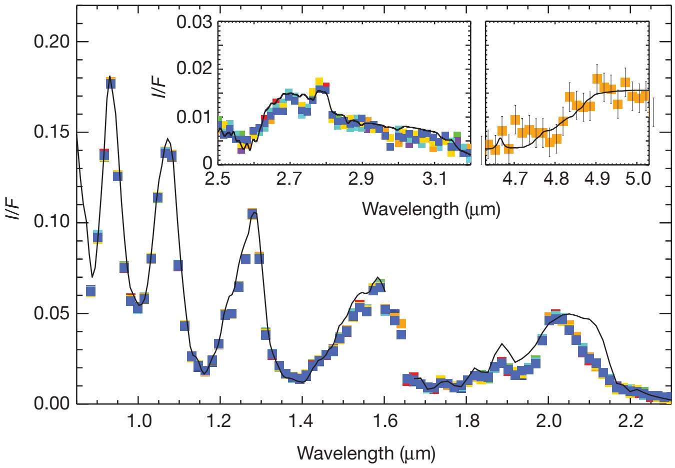


Figure 3 | Ten spectra of the dark oval compared to a black surface model. Spectra of the exceptionally dark feature at 14° S and 173° W match the simulated spectrum produced from a radiative transfer model that assumes at all wavelengths a zero surface albedo. All ten spectra (VIMS cube V1567239055, Fig. 3) appear for wavelengths less than 3.3 μm. The 21 spectra have I/F values and 3σ errors of 0.1365 ± 0.0006, 0.1046 ± 0.0006, 0.0644 ± 0.0006, and 0.0474 ± 0.0009, at 1.08 μm, 1.28 μm, 1.58 μm and 2.0 μm; the differences between the spectra precisely match the noise level of the data. The noisier 4.7–5.1 μm I/F values are an average of the ten spectra, shown with 2σ error bars (right inset). Synthetic spectra were calculated with radiative transfer models, assuming the *in situ* derived haze and methane parameters[11,12], with two alterations indicated by laboratory studies[28]. The methane coefficients are adjusted slightly, by adding 0.02 (km amagat)$^{-1}$ to all values[28], except those defining the 0.93 μm window, where earlier values[29] better characterize the spectra[14]. Also, the haze phase function determined above 80 km by DISR is used for the entire atmosphere[14]. At wavelengths longer than 1.6 μm, absorption coefficients of CH_4, CH_3D and CO are calculated with line-by-line analyses of HITRAN parameters[30]. The 1.9–2.3 μm analysis includes pressure-induced absorption due to H_2–N_2, H_2–H_2, and CH_4–N_2, assuming a H_2 abundance of 0.1% (ref. 14), and methane absorption[29]. At 2.6–3.1 μm, the haze single-scattering albedo is set to 0.65, which reproduces the opacity structure[14].

season and its long lifetime argue against a rain puddle, which would evaporate quickly[6]. Given that methane lakes are inherently unstable on Titan's tropical surface, their presence points to a subsurface source of liquid methane, that is, an oasis. Methane seepage is also indicated by the Huygens' landing site's stubby channels[23], which opens up the question of whether subterranean methane played any part in the formation of Huygens' heavily eroded landing site, with its damp flood plain, bordered by 100 m ridges. It is situated in a vast dune field.

General circulation models demonstrate that long-lasting tropical lakes several metres deep must be replenished, depending on the ethane content[24,25], within a ten-thousand-year timescale[5,6]. Taken together, tropical lakes and studies of Titan's lakes suggest that, currently, subterranean liquid supplies methane to Titan's surface and atmosphere. A supply of on average 6×10^{-4} kg m^{-2} yr^{-1} is needed to explain the composition of Titan's atmosphere, because methane, the progenitor of the moon's organic species, is destroyed in 10–100 million years through solar ultraviolet photolysis[26]. More observations are needed to determine whether this 4.5-billion-year-old moon is undergoing a specific recent flourish of geological activity, because it is freezing and its orbit decaying[27].

Received 6 May 2011; accepted 26 April 2012.

1. Niemann, H. B. *et al.* The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* **438,** 779–784 (2005).
2. Stofan, E. R. *et al.* The lakes of Titan. *Nature* **445,** 61–64 (2007).
3. Lorenz, R. D. *et al.* Titan's inventory of organic surface materials. *Geophys. Res. Lett.* **35,** L02206 (2008).
4. Lorenz, R. D. *et al.* The sand seas of Titan: Cassini RADAR observations of longitudinal dunes. *Science* **312,** 724–727 (2006).
5. Rannou, P., Montmessin, F., Hourdin, F. & Lebonnois, S. The latitudinal distribution of clouds on Titan. *Science* **311,** 201–205 (2006).
6. Mitchell, J. L. The drying of Titan's dunes: Titan's methane hydrology and its impact on atmospheric circulation. *J. Geophys. Res.* **113,** E8015 (2008).
7. Schneider, T., Graves, S. D. B., Schaller, E. L. & Brown, M. E. Polar methane accumulation and rainstorms on Titan from simulations of the methane cycle. *Nature* **481,** 58–61 (2012).
8. Brown, R. H. *et al.* The Cassini Visual and Infrared Mapping Spectrometer (VIMS) investigation. *Space Sci. Rev.* **115,** 111–168 (2004).
9. Lemmon, M. T., Karkoschka, E. & Tomasko, M. Titan's rotation—surface feature observed. *Icarus* **103,** 329–332 (1993).
10. Griffith, C. A. Evidence for surface heterogeneity on Titan. *Nature* **364,** 511–514 (1993).
11. Tomasko, M. G. *et al.* A model of Titan's aerosols based on measurements made inside the atmosphere. *Planet. Space Sci.* **56,** 669–707 (2008).
12. Tomasko, M. G., Bézard, B., Doose, L., Engel, S. & Karkoschka, E. Measurements of methane absorption by the descent imager/spectral radiometer (DISR) during its descent through Titan's atmosphere. *Planet. Space Sci.* **56,** 624–647 (2008).
13. Penteado, P. F. *et al.* Latitudinal variations in Titan's methane and haze from Cassini VIMS observations. *Icarus* **206,** 352–365 (2010).
14. Griffith, C. A. *et al.* Radiative transfer analyses of Titan's tropical troposphere. *Icarus* **218,** 975–988 (2012).
15. Brown, R. H. *et al.* The identification of liquid ethane in Titan's Ontario Lacus. *Nature* **454,** 607–610 (2008).
16. Grundy, W. M., Schmitt, B. & Quirico, E. The temperature-dependent spectrum of methane ice I between 0.7 and 5 μm and opportunities for near-infrared remote thermometry. *Icarus* **155,** 486–496 (2002).
17. Hayes, A. G, *et al.* Transient surface liquid in Titan's polar regions from Cassini. *Icarus* **211,** 655–671 (2011).
18. Barnes, J. W. *et al.* Spectroscopy, morphometry, and photoclinometry of Titan's dunefields from Cassini/VIMS. *Icarus* **195,** 400–414 (2008).
19. Radebaugh, J. *et al.* Dunes on Titan observed by Cassini radar. *Icarus* **194,** 690–703 (2008).
20. Griffith, C. A. *et al.* Characterization of clouds in Titan's tropical atmosphere. *Astrophys. J.* **702,** L105–L109 (2009).
21. Schaller, E. L., Roe, H. G., Schneider, T. & Brown, M. E. Storms in the tropics of Titan. *Nature* **460,** 873–875 (2009).
22. Turtle, E. P. *et al.* Rapid and extensive surface changes near Titan's equator: evidence of April showers. *Science* **331,** 1414–1417 (2011).
23. Soderblom, L. A. *et al.* Correlations between Cassini VIMS spectra and RADAR SAR images: implications for Titan's surface composition and the character of the Huygens Probe landing site. *Planet. Space Sci.* **55,** 2025–2036 (2007).
24. Griffith, C. A. Storms, polar deposits and the methane cycle in Titan's atmosphere. *Phil. Trans R. Soc. A* **367,** 713–728 (2009).
25. Tokano, T. Impact of seas/lakes on polar meteorology of Titan: simulation by a coupled GCM-sea model. *Icarus* **204,** 619–636 (2009).
26. Yung, Y. L., Allen, M. & Pinto, J. P. Photochemistry of the atmosphere of Titan—comparison between model and observations. *Astrophys. J.*, Suppl. 55, 465–506 (1984).
27. Tobie, G., Grasset, O., Lunine, J. I., Mocquet, A. & Sotin, C. Titan's internal structure inferred from a coupled thermal-orbital model. *Icarus* **175,** 496–502 (2005).
28. de Bergh, C. *et al.* Applications of a new set of methane line parameters to the modeling of Titan's spectrum in the 1.58 micron window. *Planet. Space Sci.* **61,** 85–98 (2012).
29. Karkoschka, E. & Tomasko, M. G. Methane absorption coefficients for the jovian planets from laboratory, Huygens, and HST data. *Icarus* **205,** 674–694 (2010).
30. Rothman, L. S. *et al.* The HITRAN 2008 molecular spectroscopic database. *J. Quant. Spectrosc. Radiat. Transf.* **110,** 533–572 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Research by C.A.G., J.T., L.D., C.S. and M.G.T. are funded by NASA's Planetary Astronomy and Cassini Data Analysis programmes. J.T. was also funded by a NASA Space Grant.

Author Contributions C.A.G. supervised all work, and conducted the radiative tranfer analyses. J.M.L. worked on the analyses of surface albedos within the wavelength windows. R.H.B. worked on the surface identification and, as the VIMS Pricipal Investigator, all technical aspects regarding the VIMS observations. M.G.T., L.D. and C.S. collaborated on the radiative transfer analyses. J.T. conducted searches of the VIMS data base, using software written by P.F.P.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.A.G. (griffith@lpl.arizona.edu).

LETTER

doi:10.1038/nature11123

A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan

Fusa Miyake[1], Kentaro Nagaya[1], Kimiaki Masuda[1] & Toshio Nakamura[2]

Increases in ^{14}C concentrations in tree rings could be attributed to cosmic-ray events[1–7], as have increases in ^{10}Be and nitrate in ice cores[8,9]. The record of the past 3,000 years in the IntCal09 data set[10], which is a time series at 5-year intervals describing the ^{14}C content of trees over a period of approximately 10,000 years, shows three periods during which ^{14}C increased at a rate greater than 3‰ over 10 years. Two of these periods have been measured at high time resolution, but neither showed increases on a timescale of about 1 year (refs 11 and 12). Here we report ^{14}C measurements in annual rings of Japanese cedar trees from AD 750 to AD 820 (the remaining period), with 1- and 2-year resolution. We find a rapid increase of about 12‰ in the ^{14}C content from AD 774 to 775, which is about 20 times larger than the change attributed to ordinary solar modulation. When averaged over 10 years, the data are consistent with the decadal IntCal ^{14}C data from North American and European trees[13]. We argue that neither a solar flare nor a local supernova is likely to have been responsible.

We used two individual Japanese cedar trees (tree A and tree B). We collected two series of measurements of the ^{14}C content ($\Delta^{14}C$, see Fig. 1 legend) of tree A. The first consists of biennial measurements from AD 750 to 820. The second consists of yearly measurements from AD 774 to 780. The data for overlapping years match within measurement errors, confirming that the two series of measurements are reproducible. The measurements of ^{14}C content in tree B were collected at 1-year resolution, from AD 770 to 779. The data from tree A and tree B are consistent (reduced $\chi^2 = 1.3$, degrees of freedom d.f. = 10). These data are presented in Supplementary Information.

Figure 1a shows the variation of ^{14}C content of Tree-A (after the two series of data were combined) and Tree-B for the period AD 750–820. In our data, we observe an increase of ^{14}C content of 12‰ within 1 year (AD 774–775), followed by a decrease over several years. The significance of this increase (AD 774–775) with respect to the measurement errors is 7.2σ.

In order to compare our results with IntCal98 (ref. 13), we averaged the yearly data to obtain a series with decadal time resolution. The result is shown in Fig. 1b. In the IntCal98 data, the ^{14}C content increased by about 7.2‰ over 10 years (AD 775–785). The two series are consistent with each other within measurement errors. The event causing the increased ^{14}C content in AD 775 could not have been local, because the IntCal data were obtained from North American and European trees, whereas we used Japanese trees.

To have produced a large number of ^{14}C nuclei in the atmosphere in AD 775, the cosmic-ray intensity must have increased considerably. The decadal record of another cosmogenic nuclide, ^{10}Be, can be obtained from the layers of ice or snow from Dome Fuji in Antarctica. These data include the relevant period, and exhibit a sharp peak in the ^{10}Be flux around AD 775 (ref. 14). However, the dating of ice core layers is more ambiguous than that of tree rings. The age of a layer is determined by locating several well-known volcanic events, and matching the production rate pattern of ^{10}Be with the ^{14}C production

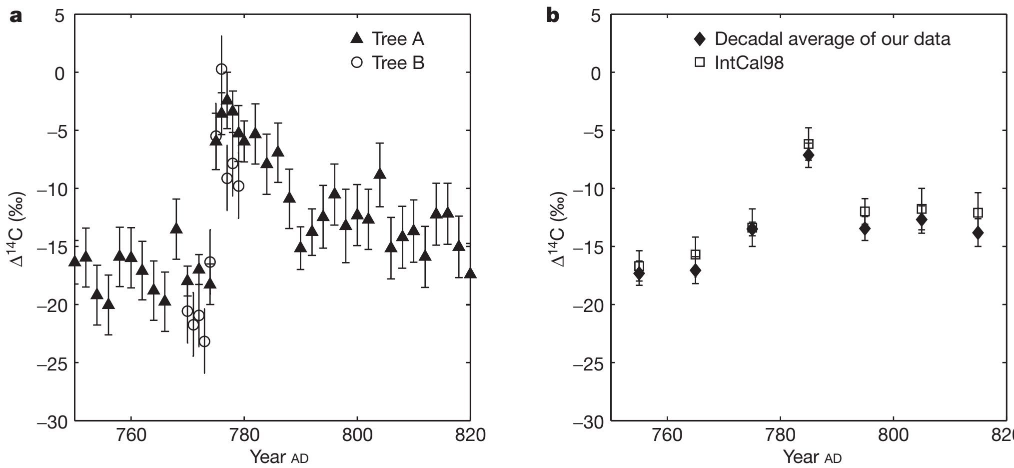


Figure 1 | Measured radiocarbon content and comparison with IntCal98. The concentration of ^{14}C is expressed as $\Delta^{14}C$, which is the deviation (in ‰) of the $^{14}C/^{12}C$ ratio of a sample with respect to modern carbon (standard sample), after correcting for the age and isotopic fractionation[30]. **a**, $\Delta^{14}C$ data for tree A (filled triangles with error bars) and tree B (open circles with error bars) for the period AD 750–820 with 1- or 2-year resolution. The typical precision of a single measurement of $\Delta^{14}C$ is 2.6‰. Most data were obtained by multiple measurements, yielding smaller errors. Error bars, 1 s.d. **b**, The decadal average of our data (filled diamonds with error bars) compared with the IntCal98 data[13] (open squares with error bars), which is a standard decadal $\Delta^{14}C$ time series. Six standard samples (NIST SRM4990C oxalic acid, the new NBS standard) were measured in the same batch of samples. Because $\Delta^{14}C$ is calculated as the deviation of the $^{14}C/^{12}C$ ratio of a sample with respect to an average of $^{14}C/^{12}C$ of the six standard samples, the errors are the resultant of error propagation. An error for a sample is a statistical one from a Poisson distribution, and an error for the standard sample is the greater of either averaged statistical error from a Poisson distribution of $\Delta^{14}C$ for the six standard samples or the s.d. of values of $^{14}C/^{12}C$ for six standard samples.

[1]Solar-Terrestrial Environment Laboratory, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan. [2]Center for Chronological Research, Nagoya University, Chikusa-ku, Nagoya 464-8601, Japan.

record reconstructed using the IntCal data. Although we cannot say with certainty that the ^{10}Be peak in the ice core occurred in AD 775, it is possible that the two peaks have the same cause. We take the agreements as further circumstantial evidence that the event was global.

To model the tree ring data, we simulated the temporal variations of ^{14}C content (using a four-box carbon cycle model) after a short-term increase in the ^{14}C production rate (based on a three-box model)[15]. Further details of this model are given in Supplementary Information. Using this model, we can calculate the hypothetical ^{14}C production rates needed to explain a rapid increase in the annual time series, for input durations of 0.1, 0.5, 1, 2 and 3 years. The best-fit values of the input ^{14}C production rate are provided in Supplementary Information. The model shows best agreement with tree ring data (Fig. 2) for a spike in ^{14}C production lasting less than 1 year. However, owing to the annual resolution of the ^{14}C data, we cannot assess the duration of this spike in more detail. Nevertheless, as the input period increases to $>$1 year, the agreement of the model with the measured data decreases. Therefore, the present data are consistent with a short-term, high-energy event producing ^{14}C, followed by a gradual decrease of ^{14}C content due to the global carbon cycle.

If the input period of ^{14}C production was 1 year, the production rate must have been 19 atoms cm^{-2} s^{-1} (see Supplementary Information) to explain the effects of this event. This is about 10 times larger than the global average production rate by galactic cosmic rays (2.05 atoms cm^{-2} s^{-1}; ref. 16).

The increment of ^{14}C content in AD 775 was about 12‰. The source cannot be the solar cycle (that is, the Schwabe cycle), which on average has an 11-year period and an amplitude of 3‰ with respect to its effect on the atmospheric ^{14}C concentration[5]. An increase of 12‰ in 1 year is about 20 times larger than expected from the Schwabe cycle. Only two known phenomena can change the cosmic-ray intensity within 1 year: a supernova explosion or a large solar proton event (SPE).

First we consider the increase of ^{14}C content due to a supernova explosion. In this case, γ-rays can produce ^{14}C because γ-rays are unaffected by the Galactic magnetic field, unlike other charged particles from supernova explosions. The production mechanism is the reaction ^{14}N(n,p)^{14}C from secondary neutrons of energy 10–40 MeV produced in the cascade from hard γ-rays in the atmosphere. No detectable increase in ^{14}C corresponding to supernovae SN 1006 and SN 1054 was reported[4], and the energy of the event in AD 775 at the Earth must be larger than these. We assume that the differential energy spectrum of γ-ray emission from a supernova is described by a power law with an index of -2.5 (ref. 4). By integrating over γ-ray energies above 10 MeV, we obtain a ^{14}C production yield of 1.2×10^2 ^{14}C atoms erg^{-1}. We computed the production yield of ^{14}C due to γ-rays using the GEANT4 simulation code with QGSP-BERT-HP[17], which is valid for thermal neutron interactions. Based on this figure, the incident γ-ray energy necessary for this increase of ^{14}C content in the atmosphere is about 7×10^{24} erg. If the distance of the supernova were the same as that of SN 1006 (2 kpc; ref. 18), the total γ-ray energy would be 3×10^{51} erg. This energy release is 100 times larger than the γ-ray energy release from a normal supernova assuming that 1% of total supernova energy goes to γ-rays and that emission of energy is isotropic (typical total supernova energy is of the order of 10^{51} erg). Therefore, the supernova was closer than 2 kpc, so that the total γ-ray energy release is 3×10^{51} erg, which is a typical supernova energy. However, although there are no historical records of a supernova visible in the Northern Hemisphere around AD 775, there are historically unrecorded supernova remnants: for example, Cassiopeia A, which was found by radio observations, or Vela Jr (RX J0852.0$-$4622), which was found by the COMPTEL γ-ray observatory, based on the ^{44}Ti line; the distance to Vela Jr is hundreds of parsecs and its age is 10^3–10^4 years (refs 19–21). Therefore, we cannot rule out an undiscovered supernova remnant corresponding to the AD 775 event. But a supernova in AD 775 may be not probable, because a supernova that occurred relatively recently and relatively near Earth should still be tremendously bright (in radio, X-rays and ^{44}Ti), and such an object is not observed.

Next we consider the case of an SPE. We assume that the flux of protons from an SPE as a function of rigidity (which is the momentum of the particle divided by the electric charge) is exponential: $\exp(-R/R_0)$, where R is the rigidity of protons and R_0 is the characteristic rigidity of the SPE. R_0 is set to 78 MV (ref. 5) in the following calculation. Unlike γ-rays, protons reaching the Earth are blocked by the geomagnetic field. We applied predicted (using EXPACS[22] software) vertical geomagnetic cut-off rigidities on the Earth for an assumed geomagnetic field the same as the present field, and calculated the flux at intervals of 10° in latitude, and obtained an average ^{14}C production yield of 10 ^{14}C atoms erg^{-1} using the GEANT4 code. The total proton energy necessary for this event was estimated to be 8×10^{25} erg at the Earth, which corresponds to 2×10^{35} erg at the Sun and may be compared to the total proton energy of 10^{29}–10^{32} erg in a normal SPE[23].

Because there is a 30% increase in the decadal ^{10}Be flux record in Dome Fuji from AD 755 to 785, we compared the production rate of ^{14}C with that of ^{10}Be (further discussions are presented in Supplementary Information.) It is possible that an SPE with an extremely hard energy spectrum could explain simultaneously the ^{14}C and ^{10}Be results, but it would have to be much harder than any flare observed so far. Furthermore, an annual time series of ^{10}Be flux would be necessary for a meticulous comparison. In fact, very large, energetic 'super flares' have been detected on normal solar-type stars. However, it is believed that a super flare has never occurred on our Sun, due to the absence of an historical record (such as a record of aurora and mass extinction caused by the expected destruction of the ozone layer[24]) and theoretical expectations[25–29].

With our present knowledge, we cannot specify the cause of this event. However, we can say that an extremely energetic event occurred around our space environment in AD 775. In the future, other high-resolution records (such as ^{10}Be and nitrate data), together with careful research of historical documentation around AD 775 and further surveys of undetected supernova remnants, may help us to clarify the cause.

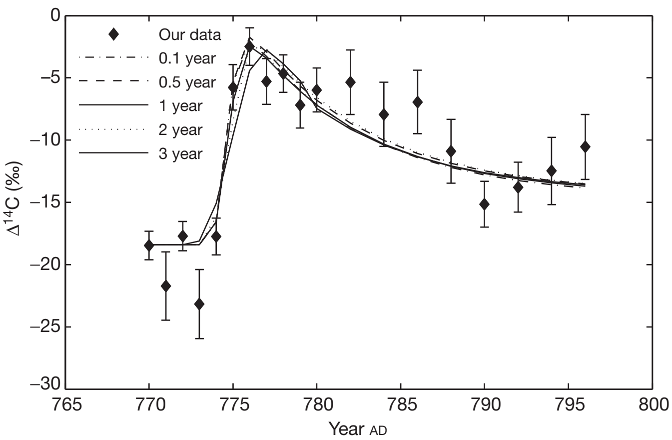


Figure 2 | Comparison of our data with a four-box carbon cycle simulation. Filled diamonds represent the Δ^{14}C values of our data, and lines represent an expected change by a four-box carbon cycle simulation. Various lines represent different cosmic-ray input durations of 0.1, 0.5, 1, 2 and 3 years. The Δ^{14}C value of the simulation in AD 773 is fixed at a value calculated by the weighted average of the three data from AD 770 to 772. Error bars, as in Fig. 1 legend.

Received 17 September 2011; accepted 4 April 2012.
Published online 3 June 2012.

1. Konstantinov, B. P. & Kocharov, G. E. *Astrophysical Events and Radiocarbon* (NASA-CR-77812, ST-CMG-AC-10430, 1965).
2. Damon, P. E., Kaimei, D., Kocharov, G. E., Mikheeva, I. B. & Peristykh, A. N. Radiocarbon production by the gamma-ray component of supernova explosions. *Radiocarbon* **37,** 599–604 (1995).

3. Damon, P. E. & Peristykh, A. N. Radiocarbon calibration and application to geophysics, solar physics, and astrophysics. *Radiocarbon* **42,** 137–150 (2000).
4. Menjo, H. *et al.* in *Proc. 29th Int. Cosmic Ray Conf.* Vol. 2 (ed. Acharya, B. S.) 357–360 (Tata Institute of Fundamental Research, Mumbai, 2005).
5. Usoskin, I. G., Solanki, S. K., Kovaltsov, G. A., Beer, J. & Kromer, B. Solar proton events in cosmogenic isotope data. *Geophys. Res. Lett.* 33**,** L08107, http://dx.doi.org/10.1029/2006GL026059 (2006).
6. Brakenridge, G. R. Core-collapse supernovae and the Younger Dryas/terminal Rancholabrean extinctions. *Icarus* **215,** 101–106 (2011).
7. LaViolette, P. A. Evidence for a solar flare cause of the Pleistocene mass extinction. *Radiocarbon* **53,** 303–323 (2011).
8. McCracken, K. G., Dreschhoff, G. A. M., Zeller, E. J., Smart, D. F. & Shea, M. A. Solar cosmic ray events for the period 1561–1994 1. Identification in polar ice, 1561–1950. *J. Geophys. Res.* **106,** 21585–21598 (2001).
9. Motizuki, Y. *et al.* An Antarctic ice core recording both supernovae and solar cycles. Preprint at http://arXiv.org/abs/0902.3446 (2009).
10. Reimer, P. J. *et al.* IntCal09 and marin09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51,** 1111–1150 (2009).
11. Stuiver, M., Reimer, P. J. & Braziunas, T. F. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* **40,** 1127–1151 (1998).
12. Takahashi, Y. *et al.* in *Proc. 30th Int. Cosmic Ray Conf.* Vol. 1 (ed. Caballero, R.) 673–676 (Universitad nacional autonoma de Mexico, 2007).
13. Stuiver, M. *et al.* INTCAL98 Radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* **40,** 1041–1083 (1998).
14. Horiuchi, K. *et al.* Ice core record of ^{10}Be over the past millennium from Dome Fuji, Antarctica: a new proxy record of past solar activity and a powerful tool for stratigraphic dating. *Quat. Geochronol.* **3,** 253–261 (2008).
15. Nakamura, T., Nakai, N. & Ohishi, S. Applications of environmental ^{14}C measured by AMS as a carbon tracer. *Nucl. Instrum. Methods B* **29,** 355–360 (1987).
16. Masarik, J. & Beer, J. An updated simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *J. Geophys. Res.* 114**,** D11103, http://dx.doi.org/10.1029/2008JD010557 (2009).
17. GEANT4. http://www.geant4.org/geant4.
18. Burrows, A. Supernova explosions in the Universe. *Nature* **403,** 727–733 (2000).
19. Iyudin, A. F. *et al.* Emission from ^{44}Ti associated with a previously unknown Galactic supernova. *Nature* **396,** 142–144 (1998).
20. Katsuda, S., Tsunemi, H. & Mori, K. Is Vela Jr. a young supernova remnant? *Adv. Space Res.* **43,** 895–899 (2009).
21. Telezhinsky, I. A new model for Vela Jr. supernova remnant. *Astropart. Phys.* **31,** 431–436 (2009).
22. Sato, T., Yasuda, H., Niita, K., Endo, A. & Sihver, L. Development of PARMA: PHITS based Analytical Radiation Model in the Atmosphere. *Radiat. Res.* **170,** 244–259 (2008).
23. Baker, D. N. in *Space Weather: The Physics Behind a Slogan* (eds Scherer, K., Fichtner, H., Heber, B. & Mall, U.) 3 (Lecture Notes in Physics, Vol. 656, Springer, 2004).
24. Schaefer, B. E., King, J. R. & Deliyannis, C. P. Superflares on ordinary solar-type stars. *Astrophys. J.* **529,** 1026–1030 (2000).
25. Lanza, A. F. Hot Jupiters and stellar magnetic activity. *Astron. Astrophys.* **487,** 1163–1170 (2008).
26. Ip, W. H., Kopp, A. & Hu, J. H. On the star-magnetosphere interaction of close-in exoplanets. *Astrophys. J.* **602,** L53–L56 (2004).
27. Willson, L. A. & Struck, C. Hot flashes on Miras? *J. Am. Assoc. Variable Star. Obs.* **30,** 23–25 (2001).
28. Struck, C., Cohanim, B. E. & Wilson, L. A. Continuous and burst-like accretion on to substellar companions in Mira winds. *Mon. Not. R. Astron. Soc.* **347,** 173–186 (2004).
29. Cuntz, M., Saar, S. H. & Musielak, Z. E. On stellar activity enhancement due to interactions with extrasolar giant planets. *Astrophys. J.* **533,** L151–L154 (2000).
30. Stuiver, M. & Polach, H. A. Discussion: reporting of ^{14}C data. *Radiocarbon* **19,** 355–363 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. Kimura for providing our tree B sample and dating the sample tree rings by dendrochronology. We also thank Y. Itow and Y. Matsubara for commenting on our manuscript. This work was partly supported by Grants-in-Aid for Scientific Research (B:22340144) provided by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

Author Contributions K.M. conducted the research. F.M. prepared samples. T.N. measured ^{14}C content by AMS at Nagoya University. F.M., K.M. and K.N. discussed the result. F.M. prepared the manuscript. K.M. and T.N. commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.M. (fmiyake@stelab.nagoya-u.ac.jp).

doi:10.1038/nature11074

'Big Bang' tomography as a new route to atomic-resolution electron tomography

Dirk Van Dyck[1] & Fu-Rong Chen[2]

Until now it has not been possible to image at atomic resolution using classical electron tomographic methods[1], except when the target is a perfectly crystalline nano-object imaged along a few zone axes[2]. The main reasons are that mechanical tilting in an electron microscope with sub-ångström precision over a very large angular range is difficult, that many real-life objects such as dielectric layers in microelectronic devices impose geometrical constraints and that many radiation-sensitive objects such as proteins limit the total electron dose. Hence, there is a need for a new tomographic scheme that is able to deduce three-dimensional information from only one or a few projections. Here we present an electron tomographic method that can be used to determine, from only one viewing direction and with sub-ångström precision, both the position of individual atoms in the plane of observation and their vertical position. The concept is based on the fact that an experimentally reconstructed exit wave[3,4] consists of the superposition of the spherical waves that have been scattered by the individual atoms of the object. Furthermore, the phase of a Fourier component of a spherical wave increases with the distance of propagation at a known 'phase speed'. If we assume that an atom is a point-like object, the relationship between the phase and the phase speed of each Fourier component is linear, and the distance between the atom and the plane of observation can therefore be determined by linear fitting. This picture has similarities with Big Bang cosmology, in which the Universe expands from a point-like origin such that the distance of any galaxy from the origin is linearly proportional to the speed at which it moves away from the origin (Hubble expansion). The proof of concept of the method has been demonstrated experimentally for graphene with a two-layer structure and it will work optimally for similar layered materials, such as boron nitride and molybdenum disulphide.

Consider a coherent plane electron wave that interacts with a single atom. If we assume the atom to be a single point, it acts as a source for a spherical wave (Ewald sphere) that propagates to the plane of detection (the image plane), where it interferes with the spherical waves emitted by the other atoms.

Using focal series reconstruction[3,4] or off-axis holography[5], it is possible to reconstruct the exit wave of the object (in the future this might even be possible using phase plates[6]). The challenge is how to determine the three-dimensional position of every individual atom of the object from the exit wave. Every spherical wave can be decomposed in terms of Fourier components. In the Fresnel approximation for the spherical wave, which is valid for high-energy electrons, the phase of each Fourier component varies linearly with increasing distance from the source and is given by $\pi\lambda g^2 f$, where λ is the wavelength, g is the spatial frequency and f is the focal distance between the atom and the plane at which the exit wave is reconstructed. Thus, if we select the exit wave around the projection of a particular atom, Fourier transform the wave and plot the respective phases of the Fourier components as function of the square of the spatial frequency, we obtain a straight line. This plot is analogous with the Hubble plot[7,8] in cosmology, which shows that the distance and recessional speed of a distant galaxy are related linearly. By linear fitting of our plot, we obtain the vertical distance from the atom to the plane of observation (the reconstructed exit wave). In our analogy, this distance is the counterpart of the time between the Big Bang[7,8] and the present (Fig. 1).

We assume that for high-energy electrons the scattering is forward and that a single atom is a weak-phase object[9]. This allows us to neglect multiple scattering and electron propagation inside the atom. Within the weak-phase object approximation, the electron wavefunction immediately behind the atom, that is, on the same side as the plane of observation, is then given by

$$\psi(r) = 1 + iV_{\mathrm{p}}(r) \qquad (1)$$

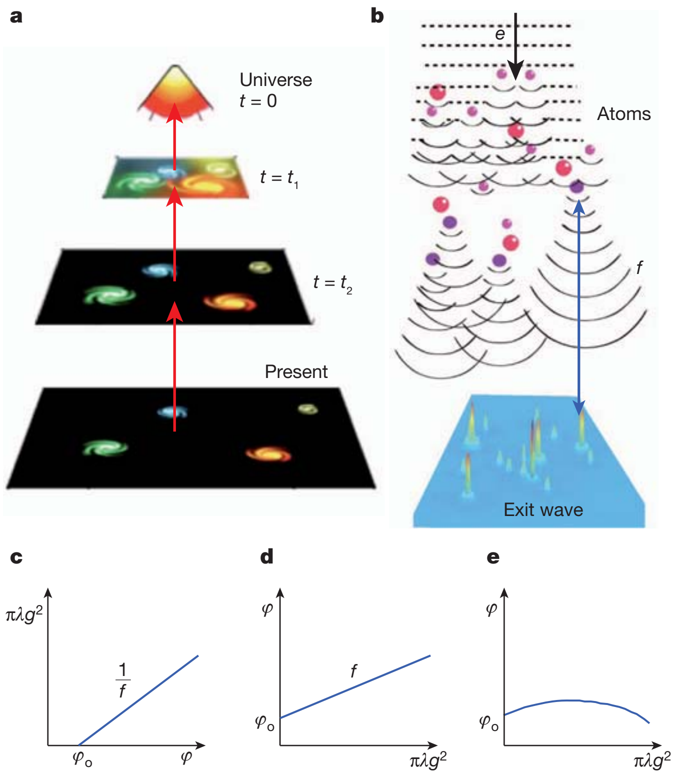


Figure 1 | Big Bang analogy. **a**, **b**, Comparison between the Big Bang (**a**) and the point-atom 'big bang' (**b**). **c**, Phase speed plotted against phase. The relationship between the two is the same as that expressed in cosmology by Hubble's law, which gives the linear relationship between the distance and the speed of a distant galaxy. Here the slope is the reciprocal focal distance, $1/f$. Note that at the position of the atom, the phase of the atom wave does not start from zero; instead, it has a value, φ_o, characteristic of the atom. **d**, Phase plotted against phase speed, which we refer to as the Hubble plot here. The slope gives the focal distance between the emitting atom and the plane of reconstruction of the exit wave. **e**, Same as in **d**, but with a minor residual spherical aberration with $C_s = 0.3$ µm (see text).

[1]University of Antwerp, Groenenborgerlaan 171, B2020 Antwerp, Belgium. [2]National Tsing Hua University, Number 101, Section 2, Kuang-Fu Road, Hsin Chu, Taiwan 300, China.

where $V_p(r)$ is the projected electrostatic potential of the atom and r is the distance to the centre of the atom in the plane of projection.

When the electron can propagate freely behind the atom, the electron wavefunction, $\psi_e(r, z)$, at a distance z measured from the centre of the atom is given in the Fresnel approximation by convolution ($\otimes$) with the Fresnel propagator $p(r, z)$, which is the parabolic approximation of the spherical wave:

$$\psi_e(r,z) = \psi(r) \otimes p(r,z) = 1 + iV_p(r) \otimes p(r,z)$$

Fourier transformation then yields

$$\psi_d(g) = F(\psi_e(r,z)) = \delta(g) + if^{el}(g)\exp(i\pi\lambda g^2 f)$$
$$= \delta(g) + f^{el}(g)\exp(i(\pi/2 + \pi\lambda g^2 f))$$

where δ is the Dirac delta function. The modulus of $\psi_d(g)$ is the scattering factor, $f^{el}(g)$, of the atom, which is the Fourier transform of the atom potential, $V_p(r)$. Because $V_p(r)$ is real and rotationally symmetric, $f^{el}(g)$ is also real and rotationally symmetric. Thus, it does not contribute to the phase of the Fourier components. Note that the factor of i produces an offset phase shift of $\pi/2$. The difference between φ, the phase at the defocus distance f, and ϕ_o, the phase at defocus 0, is more generally formulated as

$$\varphi - \varphi_o = \pi\lambda g^2 f \quad (2)$$

We can rewrite equation (2) as

$$\pi\lambda g^2 = \frac{1}{f}(\varphi - \varphi_o) \quad (3)$$

where the factor of $1/f$ is equivalent to the Hubble constant, H_o, in Hubble's law. In the weak-phase object approximation, $\varphi_o = \pi/2$.

Note also that the expression in equation (2) is rotationally symmetric. This allows us to perform rotational averaging to reduce noise, without losing any information. Figure 1c shows a theoretical 'Hubble plot' from equation (3). Unlike in the astrophysics case, we do not have to measure the phase speed, $\pi\lambda g^2$, because we know it from theory. Hence, it is more appropriate for our purposes to switch the axes and to treat the phase speed as the independent variable (Fig. 1d). The slope in Fig. 1d is the distance between the plane of the reconstructed exit wave and the emitting atom. The projected positions of the atoms in the plane of projection can be obtained with picometre precision by comparison of the phase maxima with an ideal lattice.

In practice, an exit wave can be reconstructed from high-resolution electron microscopy images using either a weighted combination of images taken at different focus values (focal series reconstruction)[3,4] or a hologram[5] obtained by interference with a reference wave. However, to maximize the resolution we must not eliminate the residual aberrations but instead properly balance them. Incoherent aberrations, such as temporal and spatial incoherence[9] and isotropic vibrations of the atoms and of the microscope, will mainly cause an isotropic blurring of the amplitude of an atom wave in real space, but not of the phase, and therefore affects only the precision[10] of the atomic position measurement. Coherent aberrations such as defocus, spherical aberration and astigmatism affect only the phase in Fourier space. The effect of residual spherical aberration, which is proportional to g^4, causes a parabolic curvature for large values of g. It turns out that our method is so sensitive that by quadratic fitting we can determine the spherical aberration constant, C_s, with a precision of more than 1 µm. Figure 1e shows a typical example of a Hubble plot with the same focal distance as in Fig. 1d but a spherical aberration with an aberration constant of $C_s = 0.3$ µm (for 80-keV electrons). As a result of this aberration, the plot deviates from a straight line. By fitting the curve with a quadratic, we can determine the residual spherical aberration constant with sub-micrometre precision. From the angular dependence of the atom wave in Fourier space, we can in principle also determine the non-symmetric higher-order residual aberrations, but for the moment we assume that these aberrations as well as the incoherent aberrations can be sufficiently corrected in the electron microscope.

We have successfully applied our method to the study of graphene using both simulations and experimentally reconstructed exit waves of single- and double-layer graphene observed with a C_s-corrected electron microscope at 80 keV. The experimental data were obtained from ref. 11. The graphene wave was reconstructed from a focal series of 19 high-resolution transmission electron microscope images[11]. The residual aberrations of the graphene exit wave were corrected up to third order by applying a numerical phase plate and by quantitative comparison with a simulated graphene exit wave[11]. Graphene is a very challenging test object for our technique because carbon atoms are very light (weak scatterers) but the distance between neighbouring atoms is very small (1.4 Å), with the result that the spherical wave of an atom is sensitive to interference from neighbours. However, the theoretical distance between the graphene layers in the double sheet is well known, and its determination therefore provides an excellent test of our method.

Figure 2 shows the phase of the exit wave of a layer of graphene that is partly overlapped by a second layer. The position analysis was carried out atom by atom, and because the theoretical positions of the atoms are known, we can estimate the statistical precision that can be obtained. There are four types of atom: those in the single layer (red); those in the lower layer that do not superpose with those in the upper layer (green); those in the two layers that superpose (brightest phase peak; blue); and those in the upper layer that do not superpose with those in the lower layer (black).

The analysis of the exit wave is done in the following steps. (i) Although the original sampling of 0.00937 nm per pixel obeys the Nyquist criterion, such that no information is lost, we need to subsample the exit wave using spline interpolation at up to 0.00268 nm per pixel (Fig. 3a) to process the data further. (ii) Figure 3b shows a subsampled area of Fig. 3a. The positions of the atoms are determined by fitting the phase peaks with Gaussian functions. The red crosses in Fig. 3b are the positions of the maxima of the fitted Gaussians. (iii) In Fig. 3c, the atom wave is isolated with a circular mark with a radius of 0.07 nm, which is half of the interatomic distance. To avoid the artefacts from the Fourier transform of a sharp circular window, we soften its edges. The background around the atom is estimated by fitting from the pixel values at the edge of the mask. (iv) A square patch of side length 0.14 nm around the isolated atom peak is selected. The modulus and phase of the isolated atom are shown colour-coded in Fig. 3d. (v) The background is calculated from the pixel values outside the mask and subtracted from the pixel values inside the mask. The modulus and phase of the isolated atom after background subtraction is shown in Fig. 3e. The value of the background wavefunction

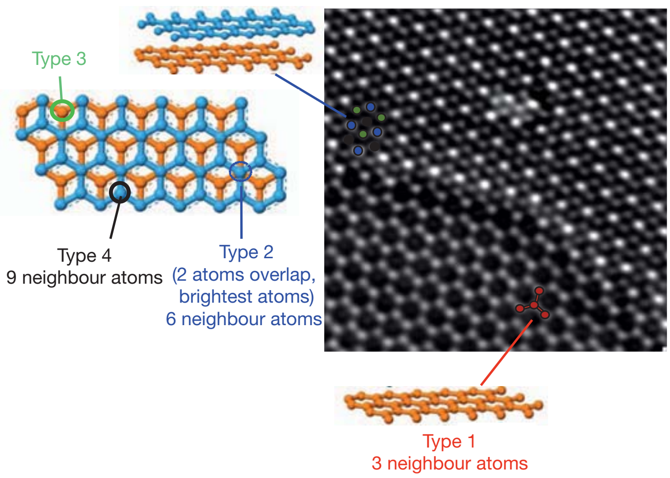


Figure 2 | Phase of the exit wave of a two-layer graphene object. Four different types of atom are distinguished: type 1 (red), type 2 (blue), type 3 (green) and type 4 (black) (see text).

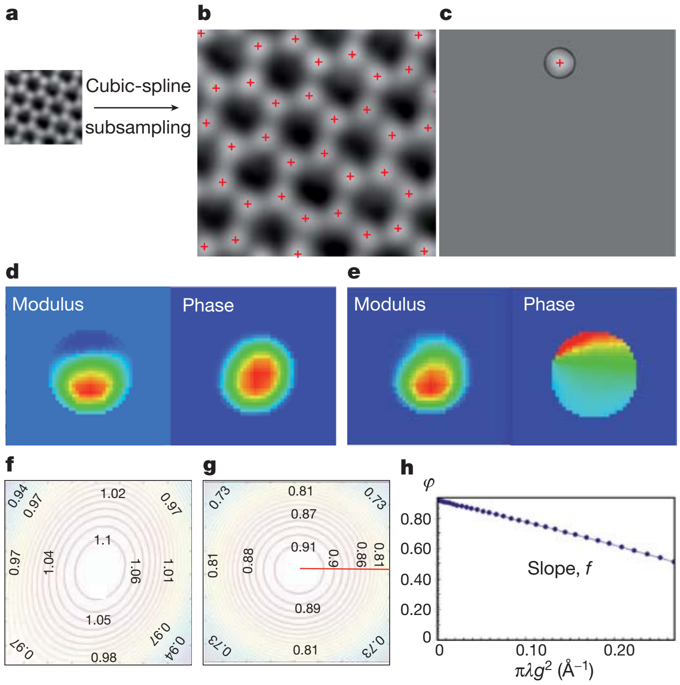


Figure 3 | Steps in the Hubble analysis. **a**, Subsampling of the exit wave. **b**, Finding the peak positions of the atom wave. **c**, Isolating the atom wave using a soft mask. **d**, Modulus and phase of the isolated atom wave. **e**, Modulus and phase of the isolated atom wave after background subtraction. **f**, Phase contour map of the Fourier transform of **e**. **g**, Phase contour map of the rotationally averaged Fourier transform. **h**, Hubble plot. The phase value is extracted from the red line in **g**.

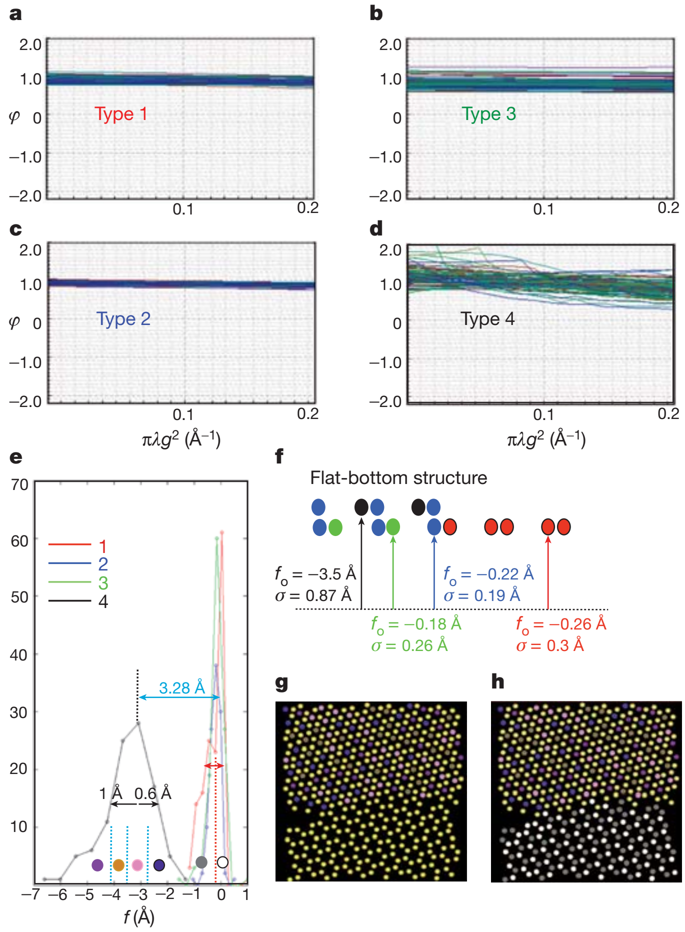


Figure 4 | Hubble plots and histogram of the focal distance. **a–d**, Hubble plots. **e**, Histogram of f for four different types of atom. **f**, The flat-bottom model. f_o, average focal distance; σ, standard deviation. **g**, Subtypes of atoms of type 4. **h**, Subtypes of atoms of type 1. Note that the sub-colours do not mean the same as in Fig. 2.

for an ideal weak-phase object is $\psi(r) = 1$ as described in equation (1). We note that after background subtraction, the modulus of the wave should maintain the shape of $V(r)$, whereas the phase should be a constant-phase plateau that is equal to $\pi/2$ for a weak-phase object. (vi) The background-subtracted data (Fig. 3e) is Fourier transformed. The phase contour map of the resulting Fourier transform is displayed in Fig. 3f. (vii) The Fourier transform of the wave (Fig. 3f) is rotationally averaged. The resulting phase contour map is depicted in Fig. 3g. (viii) The phase value φ (from the centre of the rotationally averaged wave) is plotted against phase speed, $\pi\lambda g^2$ (in Å^{-1}) (Fig. 3h). This is the Hubble plot described by equation (2). The depth of focus, f, that is, the distance between the atom and the plane of the exit wave, can be determined by linear regression of this plot.

By comparing these results with the 'correct' values given by the ideal lattice, using the method described above, we were able to determine the position of every atom in the plane of observation with an accuracy of about 7–10 pm as compared with the correct values given by the ideal graphene lattice (Supplementary Information and Supplementary Fig. 1). We were also able to determine the vertical position of every individual carbon atom to a precision of about 0.2 Å. Because every atom is measured independently, the standard deviation of the peak in the histogram of the defocus values, f (Fig. 4e), provides an internal measurement for the standard deviation of the whole fitting procedure.

Figure 4a–d shows the Hubble plots for four different types of atom in the graphene exit wave. In total, there are 143, 111, 115 and 100 atoms of types 1–4, respectively. The histogram of the distances of all the analysed atoms is given in Fig. 4e. The averaged f values for the four types of atom are -0.26, -0.22, -0.18 and -3.5 Å, respectively. This corresponds to a layer separation of 3.28 Å and shows that the graphene structure has a flat bottom. This separation is close to that of ideal two-layer graphene, which is 3.35 Å (ref. 12). As shown in Fig. 4f, the flat-bottom model shows that the atoms of types 1 (red), 2 (blue) and 3 (green) are in the bottom layer and that those of type 4 (black) are in the top layer.

To determine the quantitative limits of precision, we analysed in detail the spread in the histograms of the vertical positions. The precisions in vertical position for atoms of types 1–4 can be derived from the standard deviation of the histogram as 0.3, 0.19, 0.26 and 0.87 Å, respectively. To analyse further the origin of the larger spread for atoms of type 4, we subdivided the histogram for these atoms into four segments, each containing the same number of atoms (Fig. 4e, blue dashed lines), colour-coded magenta, brown, pink or purple according to segment, and analysed them in real space. As shown in Fig. 4g (where atoms of types 1–3 are coloured yellow), atoms of type 4 with different focal distances are distribute randomly rather than systematically. Because we have a flat-bottom structure (Fig. 4f, where the original atom colour coding is used), each atom of type 4 has nine nearest-neighbour atoms (three in the top layer and six in the bottom layer; Fig. 2). There are, however, three near neighbours for atoms of type 1 and six for atoms of types 2 and 3. It is possible that the large spread may be due to a poor signal-to-noise ratio, which may arise from the influence of nearest-neighbour atoms, because atoms of type 4 are most highly influenced in this way.

When analysing the histogram of the atoms of the single sheet (type 1), we notice a systematic difference in vertical position between two subtypes of these atoms. In Fig. 4e, the histogram for these atoms is divided at $f = -0.22$ Å into two segments (red dashed line), with reference to the positions of two subpeaks in the histogram. The atoms

were coloured grey or white if the associated focal values were less than or, respectively, greater than −0.22 Å. As shown in Fig. 4h, the difference between the average vertical positions of the white and grey atoms in the histogram suggests that the single layer may have a buckled structure, but we must still consider the possibility that this effect is due to a very small asymmetric aberration. Work on this is in progress.

If we discard the few outlying atoms of type 4 (those for which f is greater than −2.2 Å or less than −4.8 Å), the remainder have an average focal value of −3.5 Å. The three-dimensional structure deduced from our Hubble analysis of the experimental graphene exit wave can then be formed by distinguishing the outlier atoms from the rest. Normal and perspective views of this three-dimensional structure are shown in Supplementary Fig. 2b and Supplementary Fig. 2c, respectively (see Supplementary Information, Supplementary Fig. 2 and Supplementary Movie 1 for more detail).

Received 17 November 2011; accepted 22 March 2012.

1. Midgley, P. A. & Dunin-Borkowski, R. E. Electron tomography and holography in materials science. *Nature Mater.* **8,** 271–280 (2009).
2. Van Aert, S. Batenburg, K. J., Rossell, M. D., Erni, R. & Van Tendeloo, R. Three-dimensional atomic imaging of crystalline nanoparticles. *Nature* **470,** 374–377 (2011).
3. Coene, W., Thust, A., Van Dyck, D. & Op de Beeck, M. Maximum-likelihood method for focus-variation image reconstruction in high resolution transmission electron microscopy. *Ultramicroscopy* **64,** 109–135 (1996).
4. Hsieh, W. K., Chen, F. R., Kai, J. J. & Kirkland, A. I. Resolution extension and exit wave reconstruction in complex HREM. *Ultramicroscopy* **98,** 99–114 (2004).
5. Lehmann, M. & Lichte, H. Tutorial on off-axis electron holography. *Microsc. Microanal.* **8,** 447–466 (2002).
6. Van Dyck, D. Wave reconstruction in TEM using a variable phase plate. *Ultramicroscopy* **110,** 571–572 (2010).
7. Hubble, E. Effects of red shifts on the distribution of nebulae. *Astrophys. J.* **84,** 517–554 (1936).
8. Hubble, E. A relation between distance and radial velocity among extra-galactic nebulae. *Proc. Natl Acad. Sci. USA* **15,** 168–173 (1929).
9. Spence, J. C. H. *High Resolution Electron Microscopy* 3rd edn 61, 62 (Oxford Sci. Publ., 2003).
10. Bals, S., Van Aert, S., Van Tendeloo, G. & Avila-Brande, D. Statistical estimation of atomic positions from exit wave reconstruction with a precision in the picometer range. *Phys. Rev. Lett.* **96,** 096106 (2006).
11. Jinschek, J. R., Yucelen, E., Calderon, H. A. & Freitag, B. Quantitative atomic 3-D imaging of single/double sheet graphene structure. *Carbon* **49,** 556–562 (2011).
12. Reich, S., Maultzsch, J. & Thomsen, C. Tight-binding description of graphene. *Phys. Rev. B* **66,** 035412 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge discussions with A. Wang, S. Van Aert and I. Lobato. D.V.D. acknowledges financial support from the ''Research foundation - Flanders (FWO)'' under project nos G.0220.05 and G.0188.08. F.-R.C. would like to acknowledge the support from NSC-100-2120-M-007-005 and NSC-99-2120-M-007-008

Author Contributions Both authors read and commented on the paper, and contributed equally to the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.V.D. (dirk.vandyck@ua.ac.be) or F.-R.C. (frchen@ess.nthu.edu.tw).

LETTER

doi:10.1038/nature11080

Acanthodes and shark-like conditions in the last common ancestor of modern gnathostomes

Samuel P. Davis[1], John A. Finarelli[2] & Michael I. Coates[3]

Acanthodians, an exclusively Palaeozoic group of fish, are central to a renewed debate on the origin of modern gnathostomes: jawed vertebrates comprising Chondrichthyes (sharks, rays and ratfish) and Osteichthyes (bony fishes and tetrapods)[1–6]. Acanthodian internal anatomy is primarily understood from *Acanthodes bronni*[2,7–10] because it remains the only example preserved in substantial detail, central to which is an ostensibly osteichthyan braincase[1,2,7]. For this reason, *Acanthodes* has become an indispensible component in early gnathostome phylogenies[1,11–17]. Here we present a new description of the *Acanthodes* braincase, yielding new details of external and internal morphology, notably the regions surrounding and within the ear capsule and neurocranial roof. These data contribute to a new reconstruction that, unexpectedly, resembles early chondrichthyan crania. Principal coordinates analysis of a character–taxon matrix including these new data confirms this impression: *Acanthodes* is quantifiably closer to chondrichthyans than to osteichthyans. However, phylogenetic analysis places *Acanthodes* on the osteichthyan stem, as part of a well-resolved tree that also recovers acanthodians as stem chondrichthyans and stem gnathostomes. As such, perceived chondrichthyan features of the *Acanthodes* cranium represent shared primitive conditions for crown group gnathostomes. Moreover, this increasingly detailed picture of early gnathostome evolution highlights ongoing and profound anatomical reorganization of vertebrate crania after the origin of jaws but before the divergence of living clades.

Braincases are rich sources of morphological data[5], and have been used repeatedly to build hypotheses of relationships among early osteichthyans, chondrichthyans and their extinct jawed relatives: placoderms and acanthodians[1,2,7,11–13,18,19]. Notably, only a couple of acanthodian braincases have been subjected to thorough description, with only *Acanthodes bronni*[7,8,10] preserved in multiple specimens, hence the importance of obtaining an accurate understanding of this exceptional material. Existing descriptions conform the *Acanthodes* cranium to osteichthyan[7] or chondrichthyan[9] models (Supplementary Fig. 1), of which the osteichthyan version is more widely accepted[1,2,12,13]. With more early gnathostome braincases now known[11–13,15,18–21], new questions have arisen concerning apparent similarities between *Acanthodes* and its osteichthyan comparators, prompting this re-examination.

Acanthodes bronni was originally collected from Early Permian deposits (Sakmarian–Asselian, ~290–296 million years ago[22]) of Lebach, Saar-Nahe basin (southwestern Germany)[7,10], making *A. bronni* among the latest-occurring acanthodian species. Conventionally, *Acanthodes* is assigned to the Acanthodidae subdivision of the Acanthodii on the basis of its single dorsal fin, slender branchiostegals, pelvic fin proximity to pectoral fins, and absence of intermediate fin

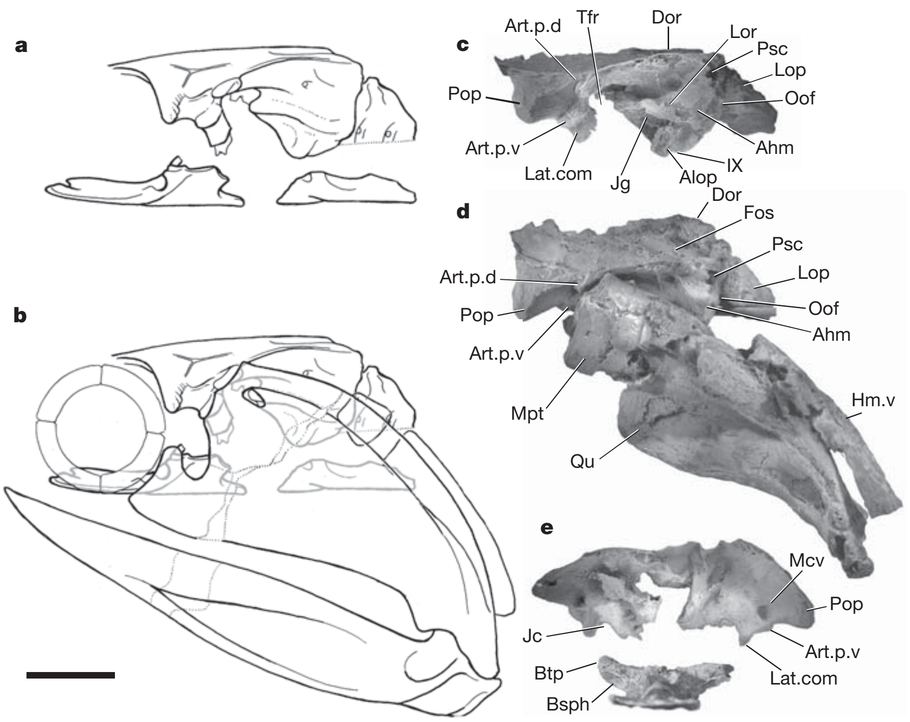


Figure 1 | *Acanthodes bronni* cranial reconstruction. **a**, **b**, Braincase (**a**) and braincase, jaws and hyomandibula (**b**) reconstructed in lateral view. **c**, Braincase dorsal ossification photographed in lateral view. **d**, Dorsal ossification articulated with palatoquadrate and hyomandibula, photographed in dorsolateral view. **e**, Dorsal ossification and basisphenoid photographed in anterior view. Ahm, hyomandibular articulation; Alop, anterolateral otic process; Art.p.d, dorsal postorbital articulation; Art.p.v, ventral postorbital articulation; Bsph, basisphenoid; Btp, basipterygoid process; Dor, dorsal ridge; Fos, fossa; Hm.v, hyomandibula ventral ossification; Jc, jugular canal; Jg, jugular groove; Lat.com, lateral commissure; Lop, lateral occipital plate; Lor, lateral otic ridge; Mcv, foramen for middle cerebral vein and anterodorsal lateral line nerve; Mpt, metapterygoid; Oof, otico-occipital fissure; Pop, postorbital process; Psc, posterior semicircular canal; Qu, quadrate; Tfr, trigemino-facial recess; IX, glossopharyngeal nerve exit. All photographs of NMS 2001.7.1 except **c**, basisphenoid from 2001.7.3. Scale bar, 10 mm.

[1]505 King Court, Smithfield, Virginia 23430, USA. [2]UCD School of Biology and Environmental Science, UCD Science Education and Research Centre, University College Dublin, Belfield, Dublin 4, Ireland. [3]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637-1508, USA.

spines[2]. Specimens (Supplementary Table 1) consist of natural moulds preserved in clay-ironstone nodules. Here, new silicone rubber casts were taken from especially well-preserved specimens of similarly sized individuals, including National Museum of Scotland (NMS) 2001.7.1 (cast of Humboldt University MB 3b) and NMS 2001.7.3 (cast of University Museum of Zoology, Cambridge GN12, formerly DMSW 495). These new data have allowed a more comprehensive reconstruction of the cranium than available in previous works (Fig. 1a, b; see Supplementary Information).

Most of the new and revised details concern the dorsal ossification of the braincase[7] (Fig. 1c–e and Supplementary Figs 2–10, 12–15), the completeness of which varies considerably among specimens. The dorsal ossification encloses most of the brain cavity, the flanking ear capsules, and the walls and processes separating the cavity and capsules from the orbits. As such, this cranial unit preserves important information regarding the organization of associated soft tissues, adding essential new information to test alternative hypotheses about the morphology[7,9] and relationships[1] of this unique taxon and the larger clade or grade of acanthodian fishes.

Differences from the standard osteichthyan model are numerous. Notably, the braincase roof bears a median dorsal ridge (Fig. 1c, d and Supplementary Figs 3a, 4) flanked by broad fossae (Fig. 1d) that terminate posteriorly above the external openings of endolymphatic ducts (Supplementary Fig. 5). This ridge–fossae–duct complex is markedly similar to conditions in early chondrichthyans[2,13,18,20]. The ridge is absent in existing descriptions of *Acanthodes*[7,9], and only parts of the endolymphatic ducts, visible on the underside of the cranial roof (Supplementary Fig. 3b), had been previously recognized[9].

Details of the orbit hind wall (Fig. 1e and Supplementary Figs 6–8) mostly agree with earlier reports, but the wide embayment in the ventral margin deserves particular attention. This has previously been identified as the trigeminofacial chamber[7], the trigeminal nerve exit[9], or the jugular canal roof[20]. We support the interpretation as the jugular canal roof (Figs 1e and 2a, b): the embayment is aligned with the jugular groove (Figs 1c and 2c, d), and the canal-space outer wall positionally matches the lateral commissure (Figs 1c, d and 2a–d) of other braincases[20,23]. The large foramen dorsolateral to this jugular canal has also been variably interpreted[7,9,20]; we support the view that it transmitted the middle cerebral vein (Figs 1e and 2a, b)[7,24].

A prominent lateral otic ridge traverses the external surface of the otic capsule (Figs 1c, 2c, d and Supplementary Fig. 9), much as in early chondrichthyans[13,18–20]. The jugular groove (Figs 1c, 2c, d and Supplementary Fig. 9) passes beneath this ridge, and below the groove projects the anterolateral otic process (Figs 1c and 2c–f). Similar

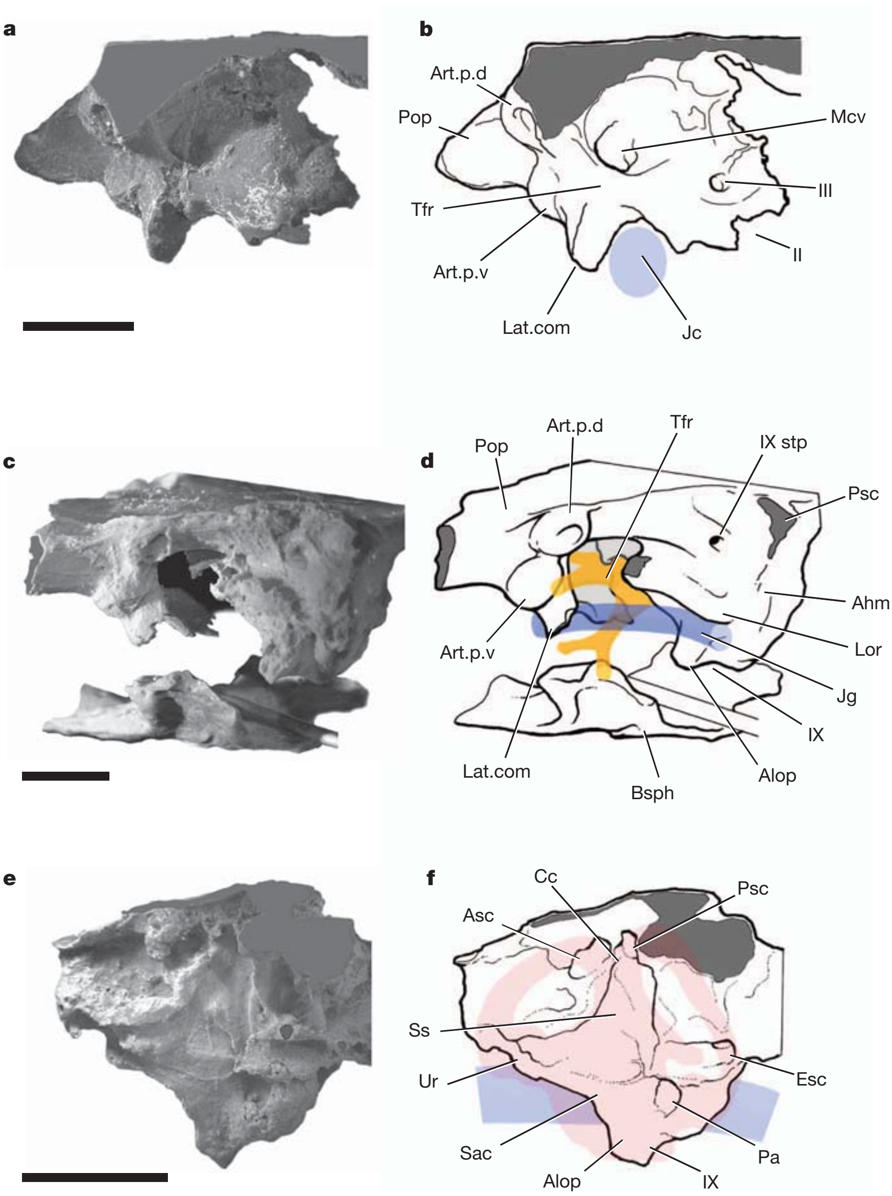


Figure 2 | *Acanthodes* braincase: details of braincase morphology. **a**, **b**, Left trigeminofacial chamber anterior wall and rear of postorbital process, posterior view. **c**, **d**, Left postorbital process, trigeminofacial recess and otic capsule, posterolateral view. **e**, **f**, Right otic capsule, medial, parasagittal view. Blue, jugular vein location; orange, trigeminal (V) and facial (VII) nerve branches; pink, semicircular canal network. Asc, anterior semicircular canal; Cc, crus commune; Esc, external semicircular canal; Pa, posterior ampulla; Sac, saccular recess; Ss, sinus superior; Ur, utricular recess; II, optic (II) nerve exit; III, oculomotor (III) nerve exit; IX stp, glossopharyngeal (IX) nerve formanen, supratemporal branch. See Fig. 1 for other abbreviations. All photographs of NMS 2001.7.1 except basisphenoid in **c** and **d**, from 2001.7.3; **a** reversed for comparison with **c**. Scale bars, 5 mm.

projections are unknown in osteichthyans or chondrichthyans, although some topological resemblance is shared with the placoderm posterior postorbital process[2,9,25]. The lateral otic ridge and anterolateral otic process are absent in previous descriptions[7,9]. Importantly, there is no evidence for a glossopharyngeal nerve foramen below the jugular groove, in contrast to osteichthyan interpretations of *Acanthodes*[2,7,12,13].

The previously unknown internal surface of the otic capsule (Fig. 2e, f) resembles early chondrichthyan and osteichthyan examples[20,23,26] (Supplementary Fig. 10). The vestibular chamber lacks a medial wall, revealing a vertical trough for the sinus superior and crus commune (Fig. 2e, f) of the inner ear canal network. However, the groove for the external semicircular canal (Fig. 2e, f) joins the base of the sinus superior dorsal to the posterior ampulla recess (Fig. 2e, f). This resembles chondrichthyans[20,27] and placoderms[9,25]; in the majority of early osteichthyans[2,9,23] the external canal and posterior ampulla are nearly, or actually, confluent (Supplementary Fig. 10b). Consistent with the absence of an external glossopharyngeal nerve foramen, there is no evidence of an internal opening, suggesting that nerve IX exited the cranium through the otico-occipital fissure (Fig. 1c, d), again resembling early chondrichthyans[19,20,26,27] (Figs 1c and 2c–f). The anterior and ventral margins of the utricular and saccular spaces (Fig. 2e, f) are uncertain because those parts of the otic wall are incomplete.

The large opening anterior to the vestibular region and behind the postorbital process contained the trigeminofacial recess[24] (Figs 1c and 2a–d), a characteristic feature of braincases in early modern gnathostomes and many extant fishes. This space maps closely to areas housing the dorsal and ventral roots and ganglia for nerves V and VII in osteichthyan and chondrichthyan crania (Supplementary Fig. 11). In *Acanthodes*, the anterior wall of this recess includes the internal opening of the canal that probably transmitted the middle cerebral vein[7] (Fig. 2a, b). As in early chondrichthyans[28], the paired articulations for the upper jaw (Figs 1c, d and 2a–d) are positioned on the rear of the postorbital process (Figs 1c–e, 2a–d and Supplementary Figs 7–9, 12). These articulations lie anterior and dorsal to the trigeminofacial recess, not on the otic capsule, as in the osteichthyan model[7]. There is no clearly defined area of articulation for the hyomandibula (Fig. 1d). We conclude that it was situated posteriorly on the lateral wall of the otic capsule (Figs 1c, d, 2c, d and Supplementary Figs 5, 9, 12).

Given the numerous morphological features shared with chondrichthyans, it might be reasoned that *Acanthodes*, and perhaps all acanthodians, are stem chondrichthyans. Principle coordinates analysis (PCO)[29,30] of a matrix of 60 taxa and 138 morphological characters (Supplementary Information) provides a means to quantify this similarity. Acanthodians, placoderms, osteichthyans and chondrichthyans form discrete clusters along the first three PCO axes (Fig. 3). Results show (1) a significant disparity between placoderms and all other gnathostomes (Supplementary Tables 2 and 3); (2) that acanthodians form a coherent group, with acanthodians more similar to one another than to other gnathostome groups (Supplementary Table 2); and (3) importantly, that *Acanthodes* in particular, and acanthodians in general, are more similar to chondrichthyans than to osteichthyans (Supplementary Tables 3–5).

In contrast to the results of these phenetic analyses, phylogenetic analysis resolves *Acanthodes* as a stem osteichthyan (Fig. 4a and Supplementary Tables 6–8), with other acanthodians distributed among osteichthyan, chondrichthyan and gnathostome stem groups. This phylogenetic result corroborates and builds upon a recent large-scale revision of the early gnathostome tree[1], delivering the most

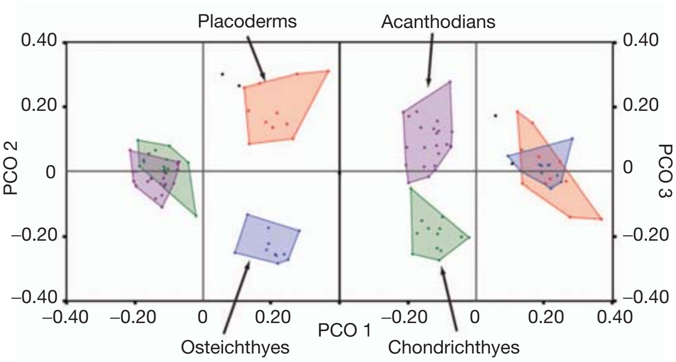


Figure 3 | PCO of early gnathostome character data. PCO 1 (17.5% explained variance) is plotted on the horizontal axis and PCO 2 (13.7%; left) and PCO 3 (10.1%; right) on the vertical axes. The four traditionally named groups (placoderms in green, acanthodians in red, osteichthyans in blue, chondrichthyans in purple) cluster in distinct and non-overlapping regions on the first three PCO dimensions. The two black points represent outgroups used in the phylogenetic analysis: Galeaspida and Osteostraci.

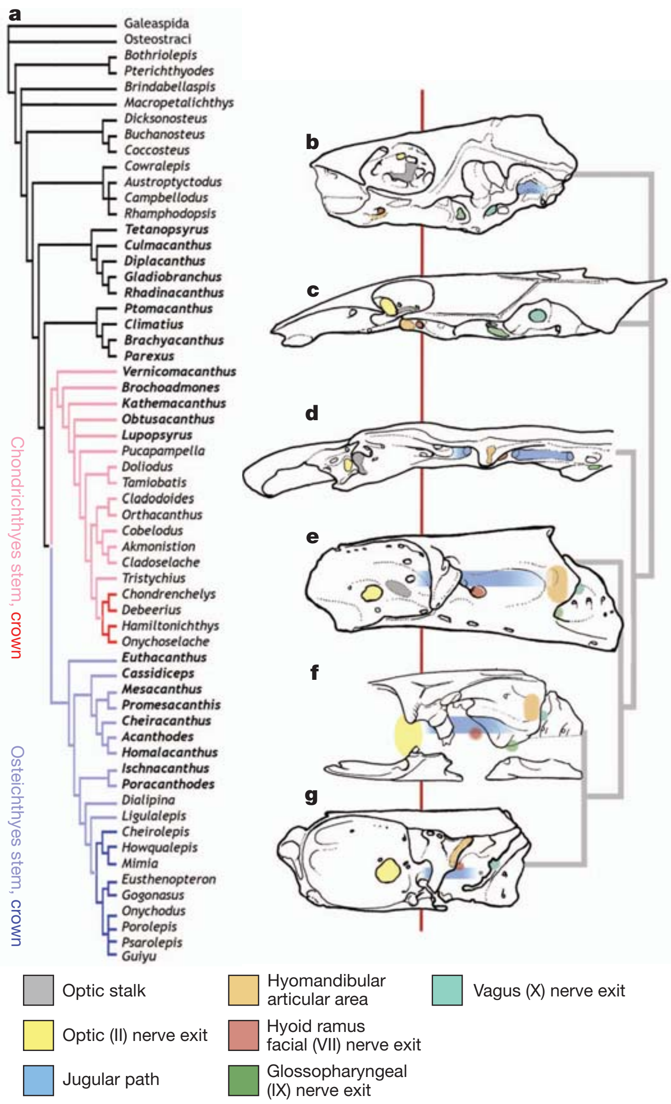


Figure 4 | Results of phylogenetic analysis, and early gnathostome braincases preceding conditions in modern jawed vertebrates. a, Strict consensus of 512 shortest cladograms; bold font signifies acanthodian genera; black branches indicate gnathostome stem group; coloured branches indicate the crown clade. **b**–**g**, Braincases in lateral view, anterior to right; simplified cladogram (grey) on left summarizes interrelationships of illustrated taxa; vertical bar (maroon) aligns braincases at level of pituitary vein canal. **b**–**d**, Placoderm-grade taxa: **b**, *Brindabellaspis*[1]; **c**, *Macropetalichthys*[9]; **d**, *Dicksonosteus*[25]. **e**–**g**, Crown group gnathostomes: **e**, *Cladodoides* (Chondrichthyes)[20]; **f**, *Acanthodes* (stem Osteichthyes); **g**, *Mimia* (crown Osteichthyes)[23] (additional and primary data sources in Supplementary Table 6).

detailed estimate of early gnathostome relationships so far. We acknowledge that tree support statistics are generally weak, and that much of the branching pattern in the crown clade hinges upon single character inclusion (Supplementary Figs 16 and 17). Moreover, character/taxon exclusion tests reveal conflicting signals within the data set and the critical influence of *Acanthodes* on recovered relationships (Supplementary Figs 18–20).

The phenetic and phylogenetic results are complementary and consistent with the hypothesis that many (perhaps all) of the chondrichthyan-like features of *Acanthodes* represent symplesiomorphies of crown group gnathostomes: compare Fig. 4e and g with f, and contrast with placoderm-grade conditions (Fig. 4b–d). The increasingly detailed estimate of early gnathostome phylogeny allows us to distinguish better among primitive and derived characteristics in chondrichthyans and osteichthyans (Supplementary Information). Character polarities illuminated by *Acanthodes* demonstrate that ancestral gnathostome conditions are not averages of living bony and cartilaginous fishes, but rather that the morphological dichotomy of modern jawed vertebrates is the result of an active diversification of osteichthyans away from an ancestral form similar to chondrichthyans. An important implication of this study is that the braincase of the last common ancestor of modern gnathostomes was effectively shark-like, of which *Acanthodes* is a narrow-based ('tropibasic') variant.

Better-resolved hypotheses of early gnathostome phylogeny are delivering a clearer picture of the ongoing re-organization of the vertebrate head, involving transformations of paired sensory organ position and size, rerouting of cranial nerves and vascular architecture, and major changes in the structural relationship between the braincase, jaws and gill arches. Such ongoing modifications started well before the origin of jaws[5], but, importantly, they continued afterwards, as is increasingly evident in the morphological disparity between members of the placoderm-grade and early crown group gnathostomes.

METHODS SUMMARY

Phylogenetic analyses were performed using the NCHUCK command to constrain searched tree space. Heuristic searches using the TBR branch-swapping option, holding one tree per replicates, were run for 25 random sequence additions to estimate shortest tree length. This estimate, plus one step, was used as the CHUCKSCORE for a further heuristic search, with 10,000 random sequence additions, keeping a maximum of 500 trees above the chuckscore per replicate. Bootstrap analyses were run for 1,000 replicates, with 100 random sequence additions per replicate, setting MAXTREES = 1,000 to constrain searched tree space. In all analyses, outgroup taxa were constrained as a paraphylum. PCO was performed on the Hamming distance matrix calculated from the character data. Further details for all analyses and all associated references are available in Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 July; accepted 22 March 2012.

1. Brazeau, M. The braincase and jaws of a Devonian 'acanthodian' and modern gnathostome origins. *Nature* **457,** 305–308 (2009).
2. Janvier, P. *Early Vertebrates* (Oxford Univ. Press, 1996).
3. Hanke, G. F. & Wilson, M. V. H. in *Morphology, Phylogeny and Paleobiogeography of Fossil Fishes* (eds Elliot, D. K., Maisey, J. G., Yu, X. & Miao, D.) 159–182 (Freidrich Pfeil, 2010).
4. Johanson, Z. Vascularization of the osteostracan and antiarch (Placodermi) pectoral fin: similarities, and implications for placoderm relationships. *Lethaia* **35,** 169–186 (2002).
5. Gai, Z., Donoghue, P. C. J., Zhu, M., Janvier, P. & Stampanoni, M. Fossil jawless fish from China foreshadows early jawed vertebrate anatomy. *Nature* **476,** 324–327 (2011).
6. Anderson, P. S. L., Friedman, M., Brazeau, M. D. & Rayfield, E. J. Initial radiation of jaws demonstrated stability despite faunal and environmental change. *Nature* **476,** 206–209 (2011).
7. Miles, R. S. in *Interrelationships of Fishes* (eds Greenwood, P. H., Miles, R. S. & Patterson, C.) 63–103 (Academic, 1973).
8. Nelson, G. J. Gill arches and the phylogeny of fishes, with notes on the classification of vertebrates. *Bull. Am. Mus. Nat. Hist.* **141,** 475–552 (1969).
9. Jarvik, E. *Basic Structure and Evolution of Vertebrates* (Academic, 1980).
10. Miles, R. S. Articulated acanthodian fishes from the Old Red Sandstone of England, with a review of the structure and evolution of the acanthodian shoulder- girdle. *Bull. Br. Mus. Nat. Hist. (Geol.)* **24,** 111–213 (1973).
11. Zhu, M. *et al.* The oldest articulated osteichthyan reveals a mosaic of gnathostome characters. *Nature* **458,** 469–474 (2009).
12. Basden, A. M., Young, G. C., Coates, M. I. & Ritchie, A. The most primitive osteichthyan braincase? *Nature* **403,** 185–188 (2000).
13. Maisey, J. G. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 263–288 (Taylor and Francis, 2001).
14. Miller, R. F., Cloutier, R. & Turner, S. The oldest articulated chondrichthyan from the Early Devonian period. *Nature* **425,** 501–504 (2003).
15. Zhu, M., Yu, X. & Janvier, P. A primitive fossil fish sheds light on the origin of bony fishes. *Nature* **397,** 607–610 (1999).
16. Coates, M. I. & Sequeira, S. E. K. in *Major Events in Early Vertebrate Evolution* (ed. Ahlberg, P. E.) 241–262 (Taylor and Francis, 2001).
17. Zhu, M., Yu, X., Wang, W., Zhao, W. & Jia, L. A primitive fish provides key characters bearing on deep osteichthyan phylogeny. *Nature* **441,** 77–80 (2006).
18. Coates, M. I. & Sequeira, S. E. K. The braincase of a primitive shark. *Trans. R. Soc. Edinb. Earth Sci.* **89,** 63–85 (1998).
19. Maisey, J. G., Miller, R. & Turner, S. The braincase of the chondrichthyan *Doliodus* from the Lower Devonian Campbellton Formation of New Brunswick, Canada. *Acta Zoologica* **90** (suppl. 1), 109–122 (2009).
20. Maisey, J. G. Braincase of the Upper Devonian shark *Cladodoides wildungensis* (Chondrichthyes, Elasmobranchii), with observations on the braincase in early chondrichthyans. *Bull. Am. Mus. Nat. Hist.* **288,** 1–103 (2005).
21. Basden, A. M. & Young, G. C. A primitive actinopterygian neurocranium from the Early Devonian of southeastern Australia. *J. Vertebr. Paleontol.* **21,** 754–766 (2001).
22. Gradstein, F. M. *et al. A Geologic Time Scale 2004* (Cambridge Univ. Press, 2004).
23. Gardiner, B. G. The relationships of the palaeoniscoid fishes, a review based on new specimens of *Mimia* and *Moythomasia* from the Upper Devonian of Western Australia. *Bull. Br. Mus. Nat. Hist. (Geol.)* **37,** 173–428 (1984).
24. Goodrich, E. S. *Studies on the Structure and Development of Vertebrates* (Univ. Chicago Press, 1930).
25. Goujet, D. *Les Poissons Placodermes du Spitsberg* (Cahiers de Paléontologie, Section Vertebres, Centre national de la Recherche scientifique, 1984).
26. Maisey, J. G. & Lane, J. A. Labyrinth morphology and the evolution of low-frequency phonoreception in elasmobranchs. *C. R. Palevol* **9,** 289–309 (2010).
27. Pradel, A. *et al.* Skull and brain of a 300 million year old chimaeroid fish revealed by synchrotron holotomography. *Proc. Natl Acad. Sci. USA* **106,** 5224–5228 (2009).
28. Maisey, J. G. The postorbital palatoquadrate articulation in elasmobranchs. *J. Morphol.* **269,** 1022–1040 (2008).
29. Wills, M. A. Crustacean disparity through the Phanerozoic: comparing morphological and stratigraphic data. *Biol. J. Linn. Soc.* **65,** 455–500 (1998).
30. Ruta, M. Phylogenetic signal and character compatibility in the appendicular skeleton of early tetrapods. *Spec. Pap. Palaeontol.* **86,** 1–21 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Paton, Z. Johanson, M. Richter, J. Clack, D. Unwin and W. Simpson for specimen loans and collections access; M. Friedman, M. Brazeau, G. Hanke and J. Long for discussions on early gnathostome cranial anatomy. Financial support for this work was provided by Natural Environment Research Council (UK) studentship GT4/97/183ES, and grant DEB-0917922 from the National Science Foundation (USA) (to M.I.C.).

Author Contributions S.P.D. completed the original data collection and initial analysis. S.P.D. and M.I.C. contributed to anatomical analysis, initiated the project and assembled the comparative data set. J.A.F. performed quantitative phenetic analyses. M.I.C. and J.A.F. contributed to phylogenetic analysis and figure preparation. All authors contributed to manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.I.C. (mcoates@uchicago.edu).

METHODS

Phylogenetic analyses of the character matrix (60 taxa coded for 138 skeletal characters: Supplementary Table 7) were performed using PAUP* v.4.0b10[31]. We report all tree lengths treating polytomies as soft. Data coverage is patchy (49.6% of the cells in the matrix are coded as uncertain/unknown) and the proportion of coded characters for individual taxa varies considerably (from 23.2% (*Rhadinacanthus*) to 89.9% (*Mimia*)). Here, we adopted tree-search strategies to efficiently explore tree space while simultaneously maximizing the probability of finding optimal islands under maximum parsimony. We used the NCHUCK command in PAUP* within heuristic searches (using the TBR branch-swapping option) to constrain the searched tree space. Heuristic searches were initially run for 25 random sequence additions to estimate the length of the shortest tree (termed TS). The value TS+1 was then used as the 'chuckscore' (the treescore at and above which a fixed number of trees were kept). A more comprehensive heuristic search was then run with 10,000 random sequence additions, keeping 500 trees greater than or equal to the chuckscore (CHUCKSCORE = TS+1, NCHUCK = 500) for each replicate[31].

We performed a parametric bootstrap[32] and calculated Bremer Decay indices[33] to assess support for the resolved nodes in the strict consensus of the most parsimonious cladograms (Supplementary Fig. 16). We performed 1,000 bootstrap replicates using heuristic searches, with 100 random sequence additions per replicate. To prevent searches from becoming stuck on large tree islands, we set the maximum number of trees saved for each random sequence addition to 1,000 (MAXTREES = 1,000). Adopting this strategy necessarily reduces the total amount of tree space searched during each random sequence addition. However, this compromise allows for both a larger number of random sequence additions per bootstrap replicate (exploring a greater breadth of tree space per random re-sampling of the character data), and a larger number of bootstrap replicates (exploring a greater number of character re-samplings). Bremer Decay indices were calculated using the Perl script AutoDecay (v.5)[34], in conjunction with PAUP*[31].

We performed additional analyses to test the sensitivity of our results to the data afforded by the new reconstruction of *Acanthodes*. We performed two taxon-subset analyses with all of the character data: (1) retaining *Acanthodes*, but removing all other taxa that have traditionally been assigned to the Acanthodii (Supplementary Fig. 18a); and (2) deleting *Acanthodes* from the analysis (Supplementary Fig. 18b). To test the sensitivity of our results to character completeness or character partitions, we performed a further pair of taxon-subset analyses. In the first, we included those taxa for which endoskeletal data are known, thereby including only the most completely coded taxa (Supplementary Fig. 19a). Second, we performed an analysis using only braincase character data (characters 1, 54–103), and including only those taxa for which the braincase is known (Supplementary Fig. 19b). In each of these supplementary phylogenetic analyses, we followed the tree-search strategy described above.

For the PCO, raw data were converted to a distance matrix that was subsequently decomposed[35] (Fig. 3). For discrete morphological character data, such as in a cladistic character matrix, intertaxon distance was measured as the number of observed character state transitions between two taxa: the Hamming distance (H_d) (for example, see ref. 36). Thus, for two taxa coded for the same 50 characters, if there is only a single character state transition (that is, a single character coded differently between the two taxa), then $H_d = 1$ (or the proportional $H_d = 0.02 = 1/50$). We normalized H_d, dividing observed character state transitions between each pair of taxa by the total number of coded characters shared between that pair, to obtain a proportional inter taxon difference ranging from 0 to 1. That is, in the case of two fully coded taxa, if there were 10 observed character state transitions between the two, then the H_d would be 10/138 = 0.072. None of the taxa in this analysis are fully coded, so only those characters coded for both were considered for normalizing H_d. In addition, we treated each character as equally weighted in the PCO analysis and did not impose character state ordering, as neither was used in the cladistic analysis.

The H_d matrix was subjected to statistical analyses to quantify the degree of similarity between the four major traditionally named gnathostome groups: placoderms, acanthodians, osteichthyans and chondrichthyans. We calculated the set of all within-group and all between-group intertaxon H_d values for each of the four traditionally named groups. To test hypotheses that the average H_d between taxa within any given named group was the same as the average distance across named groups, we performed *t*-tests[37] comparing within-group H_d for each group to its complement of three between-group sets, using Bonferroni correction for multiple comparisons[37]. To test the hypothesis that average between-group distance is the same for each potential pairing of the traditionally named groups, we performed an analysis of variance[37] across these six sets of between-group H_d values (excluding within-group distances). Post hoc tests for the significance of individual pairwise between-group H_d values were performed, using the modified T-method to account for unequal sample sizes[37]. For results of Phenetic analyses, see Fig. 3 and Supplementary Tables 1–4 for further discussion of methods used and commentary on results.

Equipment and settings. All photographs in Figures and Supplementary Figures, except for those in Supplementary Figs 2, 10 and 12 were taken using a Leica DFC490 camera attached to a Zeiss Stemi SV6 microscope. Photographs were processed using Image-Pro Plus 6.2 software; in each instance a multiple image z-stack was created, aligned and processed using the enhanced depth of field option.

31. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)* v.4.0b10 for PC (Sinauer Associates, 2002).
32. Felsenstein, J. Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* **39,** 783–791 (1985).
33. Bremer, K. The limits of amino-acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42,** 795–803 (1988).
34. Eriksson, T. *AutoDecay* v. 5.0 (2001).
35. Davis, J. C. *Statistics and Data Analysis in Geology* (John Wiley and Sons, 1986).
36. Creanza, N., Schwarz, J. S. & Cohen, J. E. Intraseasonal dynamics and dominant sequences in H3N2 influenza. *PLoS ONE* **5,** e8544 (2010).
37. Sokal, R. R., &. Rohlf. F. J. *Biometry* (W. H. Freeman, 1995).

doi:10.1038/nature11078

Covert skill learning in a cortical–basal ganglia circuit

Jonathan D. Charlesworth[1], Timothy L. Warren[1] & Michael S. Brainard[1]

We learn complex skills such as speech and dance through a gradual process of trial and error. Cortical-basal ganglia circuits have an important yet unresolved function in this trial-and-error skill learning[1]; influential 'actor–critic' models propose that basal ganglia circuits generate a variety of behaviours during training and learn to implement the successful behaviours in their repertoire[2,3]. Here we show that the anterior forebrain pathway (AFP), a cortical-basal ganglia circuit[4], contributes to skill learning even when it does not contribute to such 'exploratory' variation in behavioural performance during training. Blocking the output of the AFP while training Bengalese finches to modify their songs prevented the gradual improvement that normally occurs in this complex skill during training. However, unblocking the output of the AFP after training caused an immediate transition from naive performance to excellent performance, indicating that the AFP covertly gained the ability to implement learned skill performance without contributing to skill practice. In contrast, inactivating the output nucleus of the AFP during training completely prevented learning, indicating that learning requires activity within the AFP during training. Our results suggest a revised model of skill learning: basal ganglia circuits can monitor the consequences of behavioural variation produced by other brain regions and then direct those brain regions to implement more successful behaviours. The ability of the AFP to identify successful performances generated by other brain regions indicates that basal ganglia circuits receive a detailed efference copy of premotor activity in those regions. The capacity of the AFP to implement successful performances that were initially produced by other brain regions indicates precise functional connections between basal ganglia circuits and the motor regions that directly control performance.

We assessed the contributions of basal ganglia circuitry to learned modification of adult Bengalese finch song, a complex behaviour consisting of a sequence of 30–100-ms 'syllables', each with a highly stereotyped acoustic structure. The song-specific motor control system consists of a motor pathway, which is analogous to mammalian premotor and primary motor cortex and is sufficient to produce well-learned elements of song, and the AFP, which is necessary for juvenile song learning and adult song modification[4]. We elicited learning by training birds with aversive reinforcement contingent on the fundamental frequency of individually targeted syllables (Fig. 1a, b). Aversive reinforcement consisted of loud, 50–80-ms bursts of white noise[5,6]. Training with aversive reinforcement caused songbirds to modify fundamental frequency in a direction that adaptively reduced the likelihood of white noise exposure; delivering white noise to performances of a syllable with fundamental frequency below a threshold caused an increase in mean fundamental frequency of that syllable (Fig. 1b), whereas delivery of white noise to performances with fundamental frequency above that threshold caused a decrease in mean fundamental frequency. These adaptive changes developed within hours and were specific to the fundamental frequency of the targeted syllable.

Influential actor–critic models[2,3], inspired by reinforcement learning theory[7] and supported by empirical evidence[8,9], propose that basal ganglia circuits such as the AFP are a crucial substrate for trial-and-error learning, generating a variety of behavioural performances and ultimately implementing only the performances that have led to successful outcomes. In the context of fundamental-frequency modification (Fig. 1a, b), the actor–critic model proposes that on each trial the AFP (the actor) generates distinct fundamental frequency values (exploratory behavioural variation; Fig. 1c), receives reinforcement

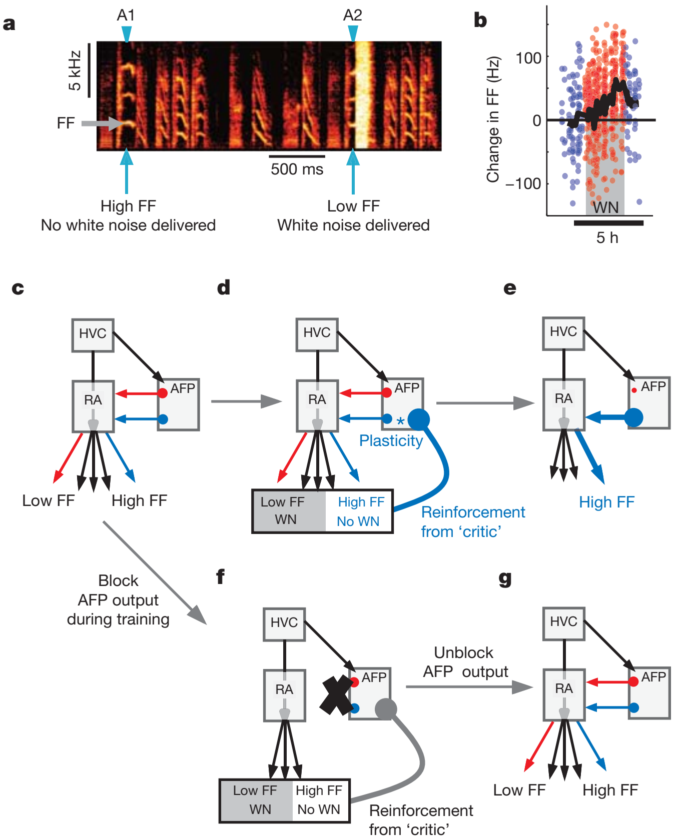


Figure 1 | Trial-and-error learning in adult birdsong. a, Spectrogram of song during an experiment in which white noise (WN) was delivered to targeted syllable (A) renditions with low fundamental frequency (FF) but not high FF. **b**, Delivering WN to syllables with low FF (shaded region) elicited increases in FF. Each point corresponds to one syllable rendition; the black line indicates the running average. **c**, The song circuit includes a motor pathway, containing HVC and RA, and the AFP, which is important for learning. The AFP generates variation in performance (motor exploration); red and light blue indicate distinct activity patterns in the AFP that lead to distinct FF values on different renditions of the same syllable. **d**, Actor–critic models propose that the AFP receives feedback about the behavioural variants that it generates, and this feedback strengthens patterns of AFP activity yielding better outcomes (light blue, feedback shown) and weakens patterns of AFP activity yielding worse outcomes (red). **e**, This changes the output of the AFP so that it selectively implements more successful behaviours. **f**, We tested this model by blocking the output of the AFP during training, thus preventing the AFP from generating variation in FF. **g**, The model predicts that this will prevent learning-related plasticity in the AFP, and thus there will be no change in FF, even when AFP output is unblocked after training.

[1]W. M. Keck Center for Integrative Neuroscience, Department of Physiology, and the Neuroscience Graduate Program, University of California, San Francisco, California 94143, USA.

signals about the consequences of that variation from dopaminergic neurons (the critic; Fig. 1d), and changes the probability of generating that fundamental frequency value in the future on the basis of its consequences[4,10–12]. Over time, the AFP gradually adjusts its output to implement (that is, to cause the execution of) behaviours with better consequences, leading to adaptive changes in fundamental frequency and thus improved skill performance (Fig. 1e). Consistent with this model, blocking AFP output through lesions or reversible inactivations reduces song variation, indicating that the AFP generates variation in song performance that might serve as motor exploration[4,5] (Fig. 1c, f). Moreover, blocking AFP output after learning reduces the expression of recently learned song changes, suggesting that the AFP can contribute to learning by biasing the motor pathway to implement more successful behaviours[13,14] (as suggested in Fig. 1e). A critical yet untested proposition of this model is that learning requires the reinforcement of exploratory behavioural variation generated by the AFP; therefore, preventing the AFP from contributing to behavioural variation during training should prevent trial-and-error learning (Fig. 1f, g).

We tested this prediction by pharmacologically blocking the output of the AFP, training birds with aversive reinforcement, and then unblocking the output of the AFP. To block contributions of the AFP to exploratory variation in song during training, while leaving intrinsic AFP circuitry intact, we exploited a pharmacological distinction between inputs that the songbird motor cortical nucleus RA (robust nucleus of the arcopallium) receives from premotor cortical nucleus HVC and from AFP output nucleus LMAN (lateral magnocellular nucleus of the anterior nidopallium). Inputs from LMAN are mediated almost exclusively by *N*-methyl-D-aspartate (NMDA) receptors whereas inputs from HVC are mediated by both NMDA receptors and α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) receptors[4] (Fig. 2a). To disrupt AFP output reversibly we therefore inserted microdialysis probes into RA and used retrodialysis to switch between a control solution (artificial cerebrospinal fluid; ACSF) and a solution containing the NMDA receptor antagonist 2-amino-5-phosphonovaleric acid (APV) at 1–5 mM (Fig. 2a). Consistent with previous reports[14,15], this manipulation affected song in the same manner as pharmacological inactivations or lesions of LMAN[14,16], reducing the coefficient of variation (CV) of the fundamental frequency by 31.7 ± 5.6% ($n = 12$ syllables in 9 birds) without causing systematic changes in song structure (Fig. 2b, c and Supplementary Fig. 2). The APV-dependent reduction in song variation was reversible; switching the infusion solution back to ACSF restored the CV of the fundamental frequency to 96.5 ± 4.6% of baseline (Fig. 2c and Supplementary Fig. 2c). These data indicate that infusing APV into RA effectively and reversibly prevents the AFP from contributing to song variation (as shown schematically in Fig. 1c, f).

As predicted by an actor–critic model of AFP function, there was no expression of learning while AFP output was blocked during training. We compared learning in control experiments (an example is shown in Fig. 3a) with learning in experiments with APV in RA throughout training (an example is shown in Fig. 3c). Training consisted of administering aversive reinforcement contingent on the fundamental frequency of a targeted syllable (Fig. 1a, b). To ensure that a similar proportion of syllable renditions received aversive reinforcement across experiments despite the reduced range of variation after APV infusion, we set the threshold for avoiding white noise at roughly the baseline median fundamental frequency for each targeted syllable (see Methods). To simplify presentation, we have plotted data so that the direction of learning (that reduces white noise exposure) is always upwards. For control experiments ($n = 14$ experiments for 9 syllables in 7 birds), there was significant expression of learning during the training period; the mean shift of fundamental frequency in the adaptive direction was 33.5 Hz, corresponding to a 1.1 ± 0.35% change in fundamental frequency (Fig. 3b, left bar; $P < 0.01$, signed-rank test). In contrast, for experiments with APV in RA ($n = 21$ experiments for 12 syllables in 9 birds), there was no expression of learning during the training period (Fig. 3d, left bar); the mean shift in fundamental frequency was 5.3 Hz (a 0.20 ± 0.15% change) which was significantly less than in control conditions ($P = 0.02$, rank-sum test) and not significantly different from zero ($P = 0.15$, signed-rank test). These results indicate that infusing APV into RA eliminates any expression of learning during training and thus provide further support that this manipulation blocks AFP output.

Learned changes to song appeared immediately when AFP output was unblocked after training. If learning required the AFP to transmit song variation during training, as predicted by an actor–critic model of AFP function, then blocking AFP output during training should have prevented learning and thus unblocking AFP output after training should not have revealed any learned changes to fundamental frequency (Fig. 1f, g). Contrary to this prediction, we observed learned changes to fundamental frequency after unblocking AFP output (Fig. 3c, d). These learned changes could not be predicted by any subtle changes in fundamental frequency during training (Supplementary Fig. 3) and were specific to the fundamental frequency of the targeted syllable (Fig. 3e and Supplementary Fig. 4). The average learned change across experiments was 27.6 Hz, corresponding to a

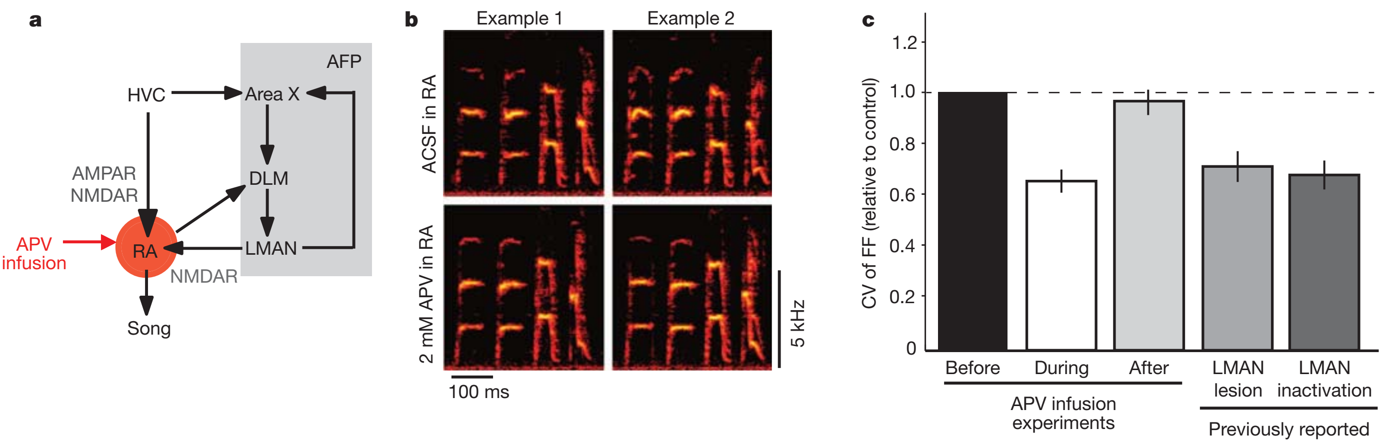


Figure 2 | Infusing APV into RA reduced song variability reversibly without distorting song structure. **a**, The AFP contains the striatopallidal nucleus Area X, the thalamic nucleus DLM and the cortical nucleus LMAN, which projects to RA. We blocked AFP output to the motor pathway by infusing the NMDA receptor (NMDAR) antagonist APV into RA. AMPAR, AMPA receptor. **b**, Infusion of APV into RA did not markedly change the song. **c**, Infusions of APV into RA reduced the coefficient of variation (CV) of FF, which recovered after switching back to ACSF ($n = 12$ syllables in 9 birds). The decrease in CV with APV in RA (31.7% ± 5.6%) was not significantly different from previously reported effects of lesions (34.1 ± 4.5%) and inactivations (28.4 ± 6.0%) of LMAN in adult Bengalese finches. Error bars indicate s.e.m. Previously reported values are from refs 14 and 16.

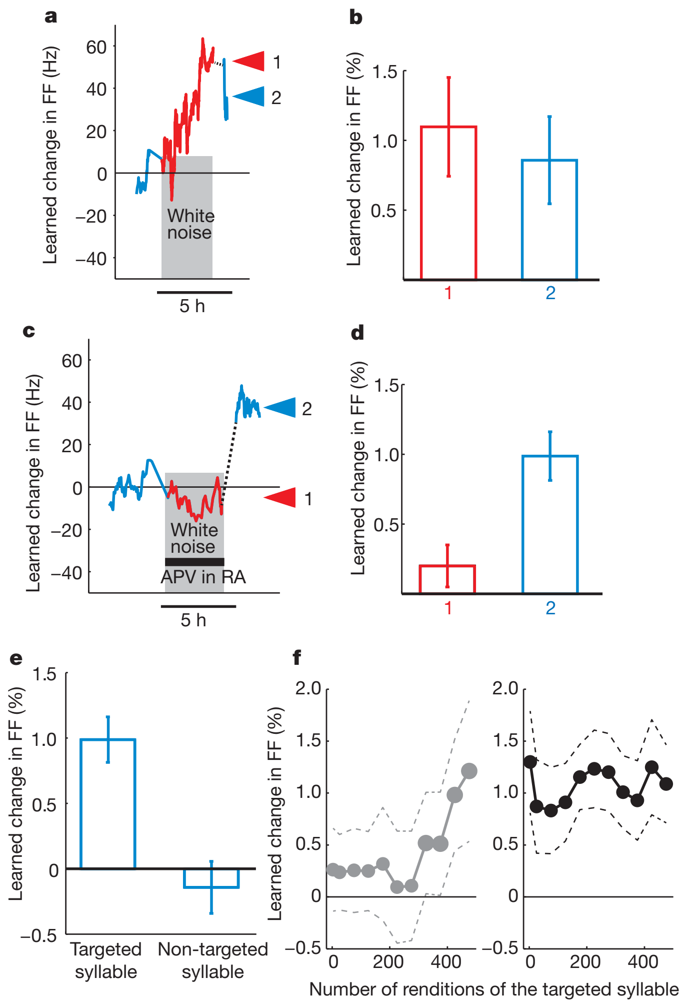


Figure 3 | Infusing APV into RA prevents the expression but not the acquisition of learning. **a**, **b**, Control experiments (ACSF in RA). **a**, Example of experiment in which white noise was delivered to targeted syllables with low FF. Arrowheads indicate FF at end of training (1) and after training (2). The dashed line indicates the delay between measurements at the end of training and after training. **b**, For control experiments ($n = 14$ experiments in 7 birds), learning was expressed at a similar magnitude at the end of training (1) and after training (2). Learning was normalized as a percentage of baseline FF. Error bars indicate s.e.m. **c**, **d**, Experiments with APV infused into RA. **c**, Example of experiment with AFP output blocked throughout the training period. Arrowheads indicate FF at end of training (1) and after training and APV washout (2). **d**, For experiments with APV in RA ($n = 21$ experiments in 9 birds), learning at end of training (1) was not significantly greater than zero and was significantly less than in control experiments. Learning after training and APV washout (2) was significantly greater than zero and was the same magnitude as in control experiments. **e**, After training and APV washout, learning was evident in syllables targeted with reinforcement (left) but not in other syllables of the same songs that were not targeted with reinforcement (right). This analysis was performed for each experiment in which FF of a non-targeted syllable could be reliably quantified ($n = 17$ of 21 total experiments). **f**, Mean progression of learning for control experiments (left) and after unblocking AFP output for experiments with APV in RA (right). Points correspond to syllable renditions 1–5, 1–50, 51–100, …, 451–500. Dashed lines indicate s.e.m.

$0.99 \pm 0.17\%$ change in fundamental frequency ($n = 21$ experiments in 9 birds; Fig. 3d, right bar; $P < 0.001$, signed-rank test). The magnitude of learning expressed after training was statistically indistinguishable from the magnitude of learning in control experiments (Fig. 3b, d, right bars; $P > 0.9$, rank-sum test). In contrast to the gradual progression of learning in control experiments, maximal learning was expressed immediately after unblocking AFP output and did not require further practice with AFP output unblocked (Fig. 3f). Thus, during training with AFP output blocked, the AFP had not only encoded a 'policy' specifying the change in song that would improve outcomes (for example, the fundamental frequency of the targeted syllable should be increased) but had already altered its activity to implement that change.

The acquisition of learning during training with APV in RA is consistent with three classes of mechanism. First, learning could require activity in the AFP during training. Second, learning could require plasticity upstream of the AFP, possibly in the ventral tegmental area, and the AFP could merely serve as a conduit between the site of plasticity and behavioural output. Third, learning could require plasticity downstream of the AFP, in RA, but the expression of that learning could be gated by AFP output[14]. To discriminate between these possible mechanisms we inactivated LMAN during training, by infusing muscimol ($n = 12$ experiments in 3 birds) or lidocaine ($n = 2$ experiments in 1 bird) into LMAN (Fig. 4a). Whereas infusing APV into RA blocked AFP output while leaving activity in the AFP intact, inactivating LMAN not only blocked AFP output but also disrupted activity within the AFP.

We found that activity in LMAN during training is crucial for learning. Inactivating LMAN reversibly reduced variation in fundamental frequency by the same amount as lesions of LMAN or infusion of APV into RA (CV decrease of $31.2 \pm 6.5\%$, $n = 14$; Supplementary Fig. 2b). We ensured in each case that the threshold for reinforcement continued to provide a directed instructive signal during the training period despite the reduced range of fundamental frequency variation (as in APV experiments; see Methods)[6]. As with infusing APV into RA, inactivating LMAN prevented any expression of learning during training; expression of learning during training with LMAN inactivated was $-0.19 \pm 0.37\%$ ($n = 14$, $P = 0.9$, signed-rank test) in comparison with $0.90 \pm 0.09\%$ ($n = 14$, $P = 1.2 \times 10^{-4}$, signed-rank test) in control experiments (Fig. 4b–d). However, in contrast to experiments with APV in RA, inactivation of LMAN during training prevented any acquisition of learning as assessed after the washout of drug ($-0.07 \pm 0.21\%$, $n = 14$, $P = 0.95$, signed-rank test; Fig. 4c, d). These results demonstrate that inactivating AFP nucleus LMAN during training prevents the acquisition of learning, and therefore that activity within the AFP during training is essential for learning.

Taken together, our results indicate that the capacity to adaptively modify a complex motor skill developed within the AFP during training with AFP output blocked. The prevention of learning by inactivating LMAN during training indicates that activity in the AFP is required for learning (Fig. 4). The immediate transition from naive performance to learned performance when we unblocked AFP output after training (Fig. 3) demonstrates that, during training, the AFP had gained the ability to improve behaviour even though that improvement was not yet expressed. For simpler forms of conditioning[17,18], such covert learning, indicating learning-related plasticity in the brain that is not accompanied by behavioural improvement, would only require that the brain region involved in learning received coarse signals about actions and stimuli[19]. In contrast, our results indicate that the brain region involved in learning, the AFP, receives detailed information (an efference copy[20]) about the precise dynamics and timing of behavioural performance from the other brain regions controlling that performance.

Our results motivate a revision to models of song plasticity[10–12] and influential actor–critic models of skill learning[2,3], which propose that essential learning-related signals develop only in brain regions that are 'acting' (that is, controlling behaviour). In contrast, our results indicate that the essential learning-related signals necessary to adaptively bias behaviour can develop in a basal ganglia circuit, the AFP, while it is prevented from contributing to behavioural performance and motor exploration. This indicates that motor exploration (that is, variation) generated by the AFP is not necessary for learning, and therefore a source of variation independent of the AFP can be exploited for

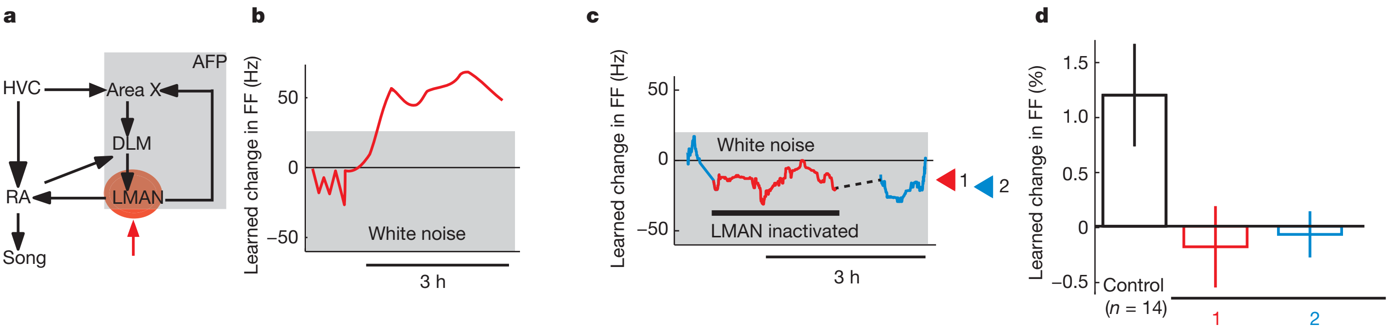


Figure 4 | Inactivating LMAN during training prevents both the expression and the acquisition of learning. **a**, We inactivated LMAN by infusing the $GABA_A$ antagonist muscimol ($n = 12$ experiments in 3 birds) or the sodium channel blocker lidocaine ($n = 2$ experiments in 1 bird) into LMAN (red arrow). **b**, **c**, Example experiments. **b**, Control experiment in which white noise was delivered to renditions of a targeted syllable with low FF. **c**, As in **b**, but with LMAN inactivated during training. Arrowheads indicate FF at the end of training with LMAN inactivated (1) and after training and muscimol washout (2). **d**, Summary: for experiments with LMAN inactivated (1 and 2; $n = 14$), there was neither evidence for learning at the end of training (red) nor after training and drug washout (light blue). Error bars indicate s.e.m.

reinforcement learning. Presumably, this variation arises in the motor pathway, possibly in RA[21,22], and is transmitted to the AFP. In normal circumstances with AFP output intact, variation contributed by the AFP itself may also be used for reinforcement learning. The AFP may therefore be a specialized hub where information about behavioural variation from multiple sources converges and is associated with reinforcement signals to guide learning.

The specificity of learning with AFP output blocked (Fig. 3e and Supplementary Fig. 4) implies that the AFP associates reinforcement signals with detailed information about ongoing song performance, including both the identity of the syllable being produced and the rendition-by-rendition variation in the fundamental frequency of that syllable. Reinforcement signals, indicating the presence or absence of white noise, could be conveyed to the AFP by means of known projections from neuromodulatory nuclei such as the ventral tegmental area[4,10]. Signals encoding syllable identity are conveyed to the AFP by means of projections from nucleus HVC in the motor pathway to the striatopallidal nucleus Area X (ref. 4). In principle, auditory feedback could provide information about variation in fundamental frequency, but such auditory signals seem to be absent from the AFP during singing[23]. We therefore favour the alternative possibility, that information about fundamental frequency variation is transmitted to the AFP through an efference copy of activity in premotor regions, by way of projections from HVC to Area X and/or projections from RA to the basal ganglia-recipient thalamic nucleus DLM (dorsolateral division of the medial thalamus)[24,25] (Supplementary Fig. 1). This is consistent with a recent proposal that the transmission of efference copy signals from motor cortex (HVC and/or RA) to basal ganglia circuitry (AFP) has a fundamental function in mammalian skill learning[26].

Our results also indicate precise functional coordination between the AFP and the motor pathway. Immediately after unblocking AFP output, we observed learning that was specific to the reinforced features of song, indicating that the AFP had modified its output to direct the production of those specific features by the motor pathway. This implies not only that the AFP receives detailed information about the song performances produced by the motor pathway during training, but also that it changes its output to specifically implement the features of those performances that were reinforced. Such a capacity of the AFP to precisely monitor and modify the activity of the motor pathway indicates fine-scale functional coordination both in the projections from the motor pathway to the AFP and in the projections from the AFP back to the motor pathway. Such bi-directional coordination might be mediated by segregated functional loops between the AFP and the motor pathway, each encoding a particular feature of song, such as high fundamental frequency in a particular syllable (Supplementary Fig. 1). Under normal conditions, with AFP output intact, such functional loops could enable the AFP to amplify and bias specific behavioural features, functions that have been attributed to mammalian basal ganglia circuits[27,28]. More generally, our results suggest that precise functional coordination between motor cortex and basal ganglia circuitry is important for enabling motor skill learning.

METHODS SUMMARY

All experiments were performed on adult (more than 120 days old) male Bengalese finches (*Lonchura striata domestica*) singing undirected song. Song recording and feedback delivery were performed with software[5] that recognized a targeted syllable and delivered a 50–80-ms burst of white noise unless the fundamental frequency (FF) met an escape criterion. For experiments with APV in RA and associated controls, the threshold for escaping white noise was set near the median FF of the targeted syllable; thus, about 50% of syllable performances initially avoided white noise. We used reverse microdialysis[14] to deliver the NMDA-receptor antagonist DL-APV (1–5 mM in ACSF) to RA, and the $GABA_A$ agonist muscimol (100–500 μM) or the sodium channel blocker lidocaine (2%) to LMAN. To ensure the complete wash-in of drug, we delayed 1–2 h between drug infusion and the beginning of the training period. Immediately after training, the solution was switched back to ACSF. To ensure complete the wash-out of drug, we delayed at least 1 h between switching the solution to ACSF and measuring FF performance after training.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 October 2011; accepted 22 March 2012.
Published online 20 May 2012.

1. Hikosaka, O., Nakamura, K., Sakai, K. & Nakahara, H. Central mechanisms of motor skill learning. *Curr. Opin. Neurobiol.* **12,** 217–222 (2002).
2. Houk, J. C., Adams, J. L. & Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 249–270 (MIT Press, 1995).
3. Suri, R. E. & Schultz, W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91,** 871–890 (1999).
4. Mooney, R. Neural mechanisms for learned birdsong. *Learn. Mem.* **16,** 655–669 (2009).
5. Tumer, E. C. & Brainard, M. S. Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* **450,** 1240–1244 (2007).
6. Charlesworth, J. D., Tumer, E. C., Warren, T. L. & Brainard, M. S. Learning the microstructure of successful behavior. *Nature Neurosci.* **14,** 373–380 (2011).
7. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
8. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
9. Reynolds, J. N., Hyland, B. I. & Wickens, J. R. A cellular mechanism of reward-related learning. *Nature* **413,** 67–70 (2001).
10. Fee, M. S. & Goldberg, J. H. A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* **198,** 152–170 (2011).
11. Fiete, I. R., Fee, M. S. & Seung, H. S. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *J. Neurophysiol.* **98,** 2038–2057 (2007).
12. Doya, K. & Sejnowski, T. in *The New Cognitive Neurosciences* (ed. Gazzaniga, M.) 469–482 (MIT Press, 2000).

13. Andalman, A. S. & Fee, M. S. A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc. Natl Acad. Sci. USA* **106,** 12518–12523 (2009).
14. Warren, T. L., Tumer, E. C., Charlesworth, J. D. & Brainard, M. S. Mechanisms and time course of vocal learning and consolidation in the adult songbird. *J. Neurophysiol.* **106,** 1806–1821 (2011).
15. Olveczky, B. P., Andalman, A. S. & Fee, M. S. Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biol.* **3,** e153 (2005).
16. Hampton, C. M., Sakata, J. T. & Brainard, M. S. An avian basal ganglia-forebrain circuit contributes differentially to syllable versus sequence variability of adult Bengalese finch song. *J. Neurophysiol.* **101,** 3235–3245 (2009).
17. Krupa, D. J., Thompson, J. K. & Thompson, R. F. Localization of a memory trace in the mammalian brain. *Science* **260,** 989–991 (1993).
18. Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W. & O'Reilly, R. C. Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nature Neurosci.* **10,** 126–131 (2007).
19. Balleine, B. W. & Ostlund, S. B. Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. NY Acad. Sci.* **1104,** 147–171 (2007).
20. Crapse, T. B. & Sommer, M. A. Corollary discharge across the animal kingdom. *Nature Rev. Neurosci.* **9,** 587–600 (2008).
21. Olveczky, B. P., Otchy, T. M., Goldberg, J. H., Aronov, D. & Fee, M. S. Changes in the neural control of a complex motor sequence during learning. *J. Neurophysiol.* **106,** 386–397 (2011).
22. Sober, S. J., Wohlgemuth, M. J. & Brainard, M. S. Central contributions to acoustic variation in birdsong. *J. Neurosci.* **28,** 10370–10379 (2008).
23. Leonardo, A. Experimental test of the birdsong error-correction model. *Proc. Natl Acad. Sci. USA* **101,** 16935–16940 (2004).
24. Vates, G. E., Vicario, D. S. & Nottebohm, F. Reafferent thalamo-'cortical' loops in the song system of oscine songbirds. *J. Comp. Neurol.* **380,** 275–290 (1997).
25. Goldberg, J. H. & Fee, M. S. A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. *Nature Neurosci.* **15,** 620–627 (2012).
26. Redgrave, P. & Gurney, K. The short-latency dopamine signal: a role in discovering novel actions? *Nature Rev. Neurosci.* **7,** 967–975 (2006).
27. Turner, R. S. & Desmurget, M. Basal ganglia contributions to motor control: a vigorous tutor. *Curr. Opin. Neurobiol.* **20,** 704–716 (2010).
28. Frank, M. J. Computational models of motivated action selection in corticostriatal circuits. *Curr. Opin. Neurobiol.* **21,** 381–386 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Frank, A. Doupe, M. Stryker and D. Mets for discussion and comments on the manuscript. This work was supported by National Institutes of Health grant NIDCD R01 and National Institute of Mental Health grant P50. J.D.C. and T.L.W. were supported by National Science Foundation graduate fellowships.

Author Contributions J.D.C., T.L.W. and M.S.B. designed the experiments. J.D.C. performed the experiments with APV in RA, and T.L.W. performed the experiments with LMAN inactivations. J.D.C. analysed the data. J.D.C. prepared the manuscript, with input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.D.C. (jcharles@phy.ucsf.edu).

METHODS

Animal care. All experiments were performed on adult (more than 120 days old) male Bengalese finches (*Lonchura striata domestica*) that had been bred in our colony and housed with their parents until at least 60 days of age. During experiments, birds were housed individually in sound-attenuating chambers (Acoustic Systems) with food and water provided *ad libitum*. All song recordings were from undirected song (that is, no female was present). All procedures were performed in accordance with established protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee.

Training. The same training parameters were used for control experiments and experiments with pharmacological manipulations. Song acquisition and feedback delivery were accomplished using previously described LabView software (EvTaf[5]), which recognized a specific time (contingency time) in a targeted syllable of song based on its spectral profile. On recognition, EvTaf recorded the time and calculated the fundamental frequency (FF) during the previous 8 ms of song. If the FF met the escape criterion (that is, above or below a threshold), no disruptive feedback was delivered. Otherwise, a 50–80-ms burst of white noise was delivered starting less than 1 ms after the contingency time. The duration of white noise was constant for a given experiment. To allow quantification of FF during training, a randomly interleaved 10% of songs were allocated as catch trials and did not receive white noise.

Experiments with reversible disruption of LMAN transmission to RA by reverse microdialysis. We interfered with LMAN transmission to RA by using a previously described reverse microdialysis technique[14], in which solution diffuses into targeted brain areas across the dialysis membranes of implanted probes. RA was mapped electrophysiologically during cannula implantation so as to direct probes to the centre of RA. Between probe insertion and white noise training, there was a more than 48 h period in which control solution (ACSF) was dialysed at a flow rate of 1 μl min^{-1}. The dialysis solution was switched from ACSF to the NMDA-receptor antagonist DL-APV (2–5 mM in ACSF; Ascent) at least 1.5 h before the onset of white noise training so that the threshold for escaping white noise could be determined on the basis of song performance with APV in RA. During this period we evaluated the efficacy of APV by assessing the rendition-to-rendition variability of FF for individual syllables. FF variability reduced and stabilized at an asymptotic level within the first 30 min of APV dialysis, indicating rapid onset and equilibrium of drug effect. We observed a reduction in variability similar to that reported after lesions or inactivations of LMAN[14,16]. For clarity of presentation in Fig. 3, running averages of FF performance for experiments with APV in RA omit the period during APV wash-in before the onset of white noise. For experiments with APV in RA and the accompanying control experiments, white noise was delivered for 4–14 h while birds were awake. Blocking AFP output reduced variation in FF by an average of 31.7%, meaning that setting the threshold for avoiding white noise at a certain level above mean FF (for example +30 Hz) in control experiments and experiments with AFP output blocked would result in a greater proportion of syllable performances escaping aversive reinforcement in control experiments. To avoid this confound and ensure that a similar proportion of syllable renditions received aversive reinforcement in control experiments and experiments with AFP output blocked, we set the threshold for avoiding white noise at approximately the baseline median FF performance (between the 40th and 60th centiles in all experiments). To ensure that our assessment of learning during the training period evaluated the effects of white noise training as opposed to the acute effects of APV, FF change at the end of the training period was quantified by subtracting FF immediately before training (during the period with APV in RA before the onset of white noise) from FF at the end of the training period. Immediately after the conclusion of white noise training, the dialysis solution was switched back to ACSF. Learning after the training period was quantified by measuring the difference between FF performance after white noise training (with ACSF in RA) and FF performance before white noise training and before infusing APV into RA (that is, with ACSF in RA). Although the latency between switching the solution remotely at the pumping apparatus and changing the solution at the probe tips was only 6 min in our experimental setup[14], the APV-dependent decrease in FF variability typically remained for hours after switching back to ACSF, presumably reflecting the combined kinetics of passive diffusion, active clearance and degradation mechanisms. In all experiments, birds were prevented from singing for at least 1.5 h after being switched from APV to ACSF to provide time for APV washout. For quantification of learning expressed immediately after training (Fig. 3f), we analysed the first songs performed after this period. To further ensure that persisting effects of APV would not cause an underestimation of learning in our primary representations of the data (Fig. 3d, e), expression of learning was assessed the morning after the training period. This allowed sufficient time for the APV-dependent block of AFP output to subside while providing limited opportunity for the birds to sing in the absence of white noise; this was important because, in control conditions, singing in the absence of white noise results in a gradual loss of learned changes to fundamental frequency[5] (that is, extinction). In a subset of experiments (8 of 24), white noise training was terminated (and APV was switched to ACSF) at least 3 h before sleep. In these experiments we found that the expression of learning before sleep was significantly greater than zero (0.95 ± 0.25% change in FF, $P < 0.02$, signed-rank test) and only slightly less than learning the next morning (1.3% ± 0.18% change in FF). This indicates that washout of APV, independently of a period of sleep, was sufficient to enable the expression of learning. Probe position in RA was established by using electrophysiological mapping of RA during implantation and confirmed *post mortem* by identifying cannula tracts in brain sections stained for Nissl bodies. Additionally, in three birds, biotinylated muscimol (diluted to 500 μM; EZ-link biotin kit; Pierce) was dialysed across the diffusion membrane to estimate the path of diffusion from the membrane[14]. In these birds, probe position was determined *post mortem* by histological staining for biotin and by comparing interleaved sections stained for Nissl bodies. Spread of drug outside RA tended to be in regions dorsal to RA, along the cannula, but not into the lateral areas where nucleus Ad is located.

Experiments with reversible inactivation of LMAN by reverse microdialysis. We examined the progression of learning for data from experiments in which we transiently inactivated LMAN by using the same reverse dialysis technique that we used for infusing APV into RA[14]. To inactivate LMAN, we switched the dialysis solution from ACSF to the $GABA_A$ agonist muscimol (100–500 μM; Sigma; 3 birds, 12 experiments) or the Na^+ channel blocker lidocaine (2%; Hospira; 1 bird, 2 experiments) at a flow rate of 1 μl min^{-1}. Inactivations lasted for 3–4 h, during which a 1 μl min^{-1} flow rate was maintained. At the conclusion of inactivation, the dialysing solution was switched back to ACSF. We applied white noise contingent on FF over a total period of 2 days or more, during both control and LMAN inactivation periods. The threshold for escaping white noise was raised incrementally to drive progressive changes in FF. In each experiment, FF eventually reached a stable value because we stopped raising the threshold. We only considered LMAN inactivations on days before FF reached this stable value, to ensure that the bird retained the capacity for further learning. For each LMAN inactivation, learning after training was quantified as the difference in FF between the last 50 renditions of the syllable before infusion of drug and the first 50 renditions of the syllable after drug washout, normalized as for experiments with APV in RA. We excluded the first 1 h after switching the infusion solution to ACSF to permit washout. During the period with LMAN inactivated, which lasted a minimum of 3 h, the threshold for escaping white noise was set so that more than 10% but less than 50% of syllables escaped and thus a learning signal of differential reinforcement was present in each experiment. This is crucial for interpretation of the lack of learning in these experiments, because learning in this model does not proceed without such differential reinforcement[6]. Learning during training with LMAN inactivated was quantified with a linear regression of FF on the renditions of the targeted syllable during training with LMAN inactivated. For each inactivation, matched learning in control conditions was quantified by calculating the average hourly rate of change in FF during ACSF infusion on the day of that inactivation and multiplying that rate by the number of hours for which LMAN was inactivated. Probe positioning and the path of drug diffusion were evaluated *post mortem* by histological staining of sectioned tissue as described previously[14]. Tissue damage caused by cannulae enabled confirmation that probes were accurately targeted to LMAN. In addition, biotinylated muscimol or ibotenic acid was used to estimate the spread of diffusion, as described previously[14].

Analysis. All analyses were performed with custom software written in MATLAB (Mathworks). For a given syllable, FF was measured over a consistent time window aligned to syllable onset; for syllables targeted with white noise feedback, the measurement time window was centred on the median time at which feedback was delivered. FF was calculated as described previously[6] for both targeted syllables and non-targeted syllables of the same song. Spectral entropy, volume and duration were calculated as described previously[5]. Statistical significance was tested with non-parametric statistical tests; Wilcoxon signed-rank tests and Wilcoxon rank-sum tests were used where appropriate.

doi:10.1038/nature11015

Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2

Michael J. Schmeisser[1]*, Elodie Ey[2,3,4]*, Stephanie Wegener[5]*, Juergen Bockmann[1], A. Vanessa Stempel[5], Angelika Kuebler[1], Anna-Lena Janssen[1], Patrick T. Udvardi[1], Ehab Shiban[1]†, Christina Spilker[6], Detlef Balschun[7], Boris V. Skryabin[8,9], Susanne tom Dieck[10], Karl-Heinz Smalla[11], Dirk Montag[12], Claire S. Leblond[2,3,4], Philippe Faure[13], Nicolas Torquet[2,3,4], Anne-Marie Le Sourd[2,3,4], Roberto Toro[2,3,4], Andreas M. Grabrucker[1], Sarah A. Shoichet[5], Dietmar Schmitz[5], Michael R. Kreutz[6], Thomas Bourgeron[2,3,4], Eckart D. Gundelfinger[11] & Tobias M. Boeckers[1]

Autism spectrum disorders comprise a range of neurodevelopmental disorders characterized by deficits in social interaction and communication, and by repetitive behaviour[1]. Mutations in synaptic proteins such as neuroligins[2,3], neurexins[4], GKAPs/SAPAPs[5] and ProSAPs/Shanks[6–10] were identified in patients with autism spectrum disorder, but the causative mechanisms remain largely unknown. ProSAPs/Shanks build large homo- and heteromeric protein complexes at excitatory synapses and organize the complex protein machinery of the postsynaptic density in a laminar fashion[11,12]. Here we demonstrate that genetic deletion of ProSAP1/Shank2 results in an early, brain-region-specific upregulation of ionotropic glutamate receptors at the synapse and increased levels of ProSAP2/Shank3. Moreover, *ProSAP1/Shank2*$^{-/-}$ mutants exhibit fewer dendritic spines and show reduced basal synaptic transmission, a reduced frequency of miniature excitatory postsynaptic currents and enhanced *N*-methyl-D-aspartate receptor-mediated excitatory currents at the physiological level. Mutants are extremely hyperactive and display profound autistic-like behavioural alterations including repetitive grooming as well as abnormalities in vocal and social behaviours. By comparing the data on *ProSAP1/Shank2*$^{-/-}$ mutants with *ProSAP2/Shank3αβ*$^{-/-}$ mice, we show that different abnormalities in synaptic glutamate receptor expression can cause alterations in social interactions and communication. Accordingly, we propose that appropriate therapies for autism spectrum disorders are to be carefully matched to the underlying synaptopathic phenotype.

Many of the recently identified autism spectrum disorders (ASD) candidate genes code for proteins of excitatory synapses[13–15], suggesting that these disorders may arise from molecular imbalances of synaptic connections. In this context, targeted disruption of the *ProSAP2/Shank3* gene in mice resulted in molecular perturbations of glutamatergic synapses and profound autistic-like behaviour[16–19]. Here we generated mice lacking all isoforms of ProSAP1/Shank2 (Fig. 1A and Supplementary Fig. 1a–g) to decipher the interrelation between ProSAP1/Shank2 protein levels, synaptic architecture, neurophysiology and behaviour in mice.

Heterozygous *ProSAP1/Shank2*$^{+/-}$ (expressing approx. 50% of ProSAP1/Shank2 protein, Fig. 1A) and homozygous *ProSAP1/Shank2*$^{-/-}$ mutants were viable, but their survival rate was lower compared with wild-type littermates (Supplementary Fig. 2a). Although body weight was reduced (Supplementary Fig. 2b–e), adult mutants displayed normal appearance and overall brain morphology (data not shown). However, hindlimb clasping was observed (Supplementary Fig. 2f), similarly to some other mouse models of ASD[20,21].

Owing to high expression of ProSAP1/Shank2 in the hippocampus during spinogenesis[22] and as patient-based mutations in ProSAP1/Shank2 were recently shown to alter dendritic spines in the hippocampus[23], we assessed spine density and synaptic ultrastructure in the CA1 region. We found a small reduction of spine numbers in *ProSAP1/Shank2*$^{-/-}$ mutants (Fig. 1B, a) whereas postsynaptic density (PSD) length or thickness was not significantly altered (Fig. 1B, b). Biochemical analysis revealed higher levels of the *N*-methyl-D-aspartate receptor (NMDAR) subunit GluN1 and ProSAP2/Shank3 in whole brain PSDs of *ProSAP1/Shank2*$^{-/-}$ mice (Supplementary Fig. 3a). Interestingly, ProSAP2/Shank3 upregulation specifically occurred at synapses, as protein and messenger RNA (mRNA) levels were not changed significantly in whole brain of mutant versus wild-type animals (Supplementary Fig. 3b, c). Further evidence for local compensation was apparent by subfractionation experiments and transient knockdown of ProSAP1/Shank2 in primary hippocampal cultures, resulting in a rapid increase of GluN1 and ProSAP2/Shank3 at synaptic sites (Supplementary Fig. 3b, d).

Based on the ProSAP1/Shank2 expression profile in wild-type mouse brain (Supplementary Fig. 4), we biochemically isolated crude synaptosomal fractions from cortex, hippocampus and striatum of wild-type, *ProSAP1/Shank2*$^{+/-}$ and *ProSAP1/Shank2*$^{-/-}$ mice at postnatal day (P)25 and P70 to examine molecular alterations with respect to brain regions and development. The major change compared with wild types was an early increase of NMDAR subunits in the hippocampus and striatum of *ProSAP1/Shank2*$^{-/-}$ mutants. Notably, this increase was sensitive to *ProSAP1/Shank2* gene dosage as it was also observed in *ProSAP1/Shank2*$^{+/-}$ mice, but to a lesser extent. At P70 the upregulation of ProSAP2/Shank3 was observed in all brain regions investigated (Fig. 1C, D and Supplementary Figs 5 and 7; cortex data not shown). We compared these observations with molecular synaptic changes when major isoforms of ProSAP2/Shank3 are not present and analysed *ProSAP2/Shank3αβ*$^{-/-}$ mice (similar to recently published Shank3 mutants[17], Supplementary Fig. 6). Especially in the striatum, we observed a clear difference between *ProSAP1/Shank2*$^{-/-}$ and *ProSAP2/Shank3αβ*$^{-/-}$ mice with respect to their synaptic content of ionotropic glutamate receptors. The levels of most subunits were higher in *ProSAP1/Shank2*$^{-/-}$ and lower in *ProSAP2/Shank3αβ*$^{-/-}$ mice. Interestingly, apart from the increase of ProSAP2/Shank3 in

[1]Institute for Anatomy and Cell Biology, Ulm University, 89081 Ulm, Germany. [2]Human Genetics and Cognitive Functions, Institut Pasteur, 75724 Paris CEDEX 15, France. [3]CNRS, URA 2182 'Genes, Synapses and Cognition', Institut Pasteur, 75724 Paris CEDEX 15, France. [4]University Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, 75013 Paris, France. [5]Neuroscience Research Center, Cluster of Excellence NeuroCure, Charité, 10117 Berlin, Germany. [6]PG Neuroplasticity, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany. [7]Laboratory of Biological Psychology, Department of Psychology, Catholic University of Leuven, 3000 Leuven, Belgium. [8]Institute of Experimental Pathology (ZMBE), University of Muenster, 48149 Muenster, Germany. [9]Interdisciplinary Center for Clinical Research (IZKF), University of Muenster, 48149 Muenster, Germany. [10]Max Planck Institute for Brain Research, Department of Synaptic Plasticity, 60528 Frankfurt, Germany. [11]Department of Neurochemistry, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany. [12]Neurogenetics Special Laboratory, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany. [13]University Paris 06, CNRS, UMR 7102, 75005 Paris, France. †Present address: Klinikum rechts der Isar, Technische Universität München, Neurosurgery Department, Ismaninger Str. 22, 81675 Munich, Germany.
*These authors contributed equally to this work.

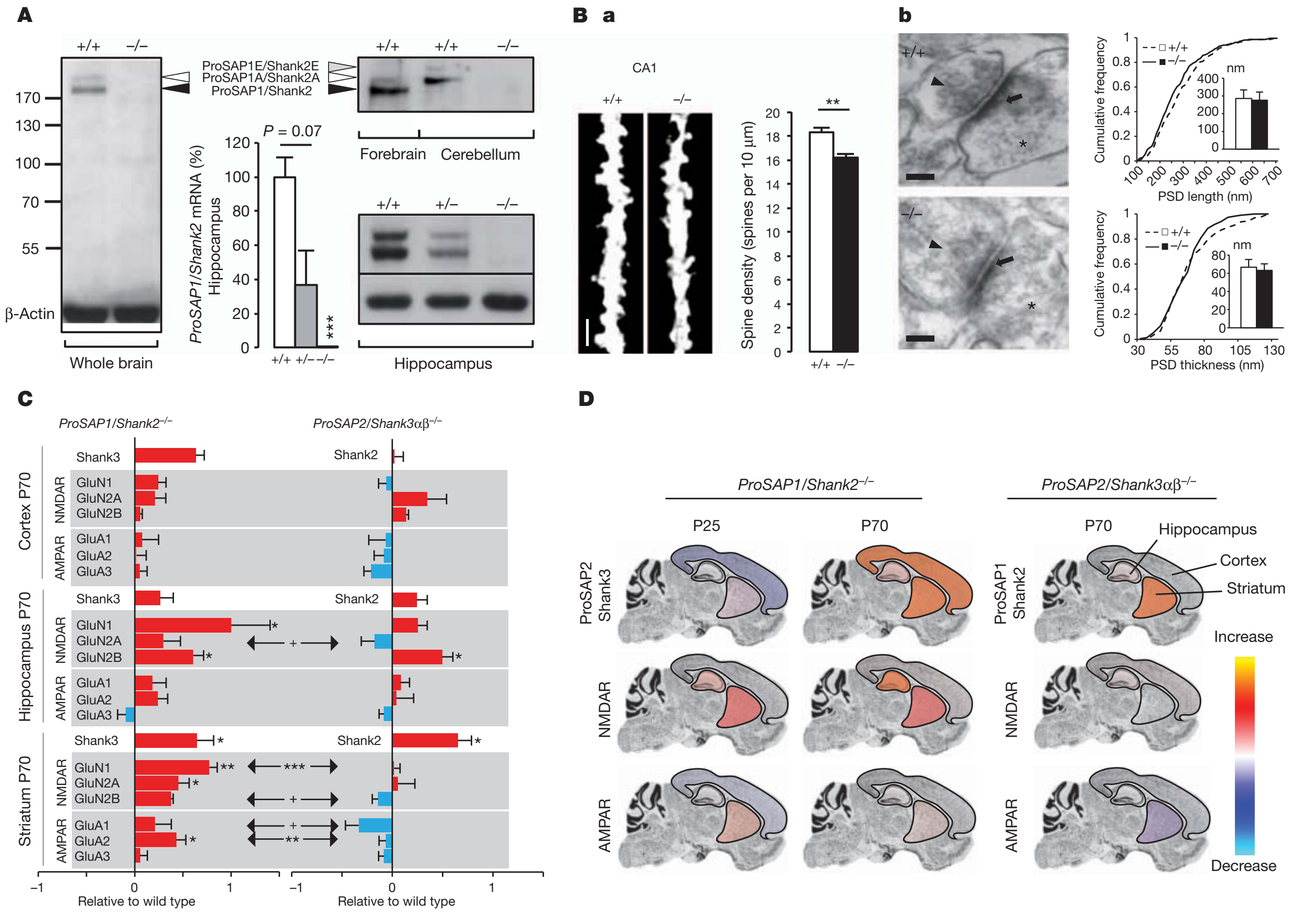


Figure 1 | Cyto-architechtural and molecular changes in *ProSAP1/Shank2*$^{-/-}$ mouse brain. **A**, Western blot of pooled ($n = 10$) whole brains (left panel) and cerebella (upper right panel) from wild-type (+/+) and *ProSAP1/Shank2*$^{-/-}$ (−/−) mice as indicated. ProSAP1/Shank2 isoforms are marked by arrowheads: ProSAP1/Shank2 (black), ProSAP1A/Shank2A (white), ProSAP1E/Shank2E (grey). Forebrain wild-type homogenate was used as control to differentiate cerebellar isoforms. Total ProSAP1/Shank2 mRNA (middle panel) and protein levels (right lower panel) from wild-type, *ProSAP1/Shank*$^{+/-}$ (+/−) and *ProSAP1/Shank2*$^{-/-}$ hippocampi. **B**, **a**, Representative images of secondary dendrites from CA1 hippocampal neurons of adult wild-type and *ProSAP1/Shank2*$^{-/-}$ mice (Golgi–Cox staining, scale bar: 1 μM) and quantification of spine density from $n = 6$ wild-type (white bar) and *ProSAP1/Shank2*$^{-/-}$ (black bar) littermate pairs. **B**, **b**, Representative electron microscopy images of CA1 synapses from wild-type and *ProSAP1/Shank2*$^{-/-}$ animals. Synaptic vesicles (arrowheads), PSDs (arrows) and dendritic spines (asterisks). Scale bar: 100 nm. Analysis of PSD length and thickness from wild-type and *ProSAP1/Shank2*$^{-/-}$ animals (right panel). Data are presented as cumulative frequency plots, small insets depict median values compared between wild-type (white bars) and *ProSAP1/Shank2*$^{-/-}$ (black bars) animals. $n = 220$ PSDs for six wild types and $n = 215$ PSDs for six *ProSAP1/Shank2*$^{-/-}$ mice. **C**, Semi-quantitative analysis of proteins in crude synaptosomal fractions from different brain regions of wild-type, *ProSAP1/Shank2*$^{-/-}$ and *ProSAP2/Shank3*$\alpha\beta^{-/-}$ mice as indicated. Mutant protein was normalized to wild-type levels and is plotted as relative change of expression levels. **D**, Colour-coded visualization of protein levels (ProSAP1/Shank2, ProSAP2/Shank3, NMDAR, AMPAR) in *ProSAP1/Shank2*$^{-/-}$ or *ProSAP2/Shank3*$\alpha\beta^{-/-}$ brains (cortex, hippocampus, striatum) at the indicated time points (P25, P70). **C**, **D**, Red bars/colour indicate elevated, blue bars/colour decreased, protein levels. **A**–**D**, +/+, wild types; +/−, *ProSAP1/Shank2*$^{+/-}$; −/−, *ProSAP1/Shank2*$^{-/-}$. All data are presented as mean ± s.e.m.; all P values are derived from unpaired, two-tailed Student's t-tests (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

ProSAP1/Shank2$^{-/-}$ mice, we detected a vice versa upregulation of ProSAP1/Shank2 in *ProSAP2/Shank3*$\alpha\beta^{-/-}$ mice (Fig. 1C, D and Supplementary Figs 5–7). This phenomenon was not due to an increase of transcript levels and was observed in whole brain PSDs from both animal models (Supplementary Fig. 8a–c).

To analyse how the altered molecular composition of *ProSAP1/Shank2*$^{-/-}$ synapses influences synaptic transmission, we performed extracellular field and whole-cell patch clamp recordings from CA1 pyramidal cells in acute hippocampal slices. Field excitatory postsynaptic potentials (fEPSPs) were decreased by approximately 40% in *ProSAP1/Shank2*$^{-/-}$ (Fig. 2A) as well as *ProSAP1/Shank2*$^{+/-}$ animals (Supplementary Fig. 9a). The reduced synaptic transmission was not only found in young mice (P21–P28), but also in older animals (3 months of age, Fig. 2A, c, investigated in *ProSAP1/Shank2*$^{-/-}$ only). There was no evidence for genotypic differences in the excitability of presynaptic fibres, the intrinsic firing threshold and the whole-cell input resistance of CA1 pyramidal cells (Supplementary Fig. 9b–d). When analysing miniature excitatory postsynaptic currents (mEPSCs) from CA1 pyramidal cells (Fig. 2B), *ProSAP1/Shank2*$^{-/-}$ mice showed a significant reduction in mEPSC frequency. There was no evidence for differences in mEPSC amplitudes and α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)-mediated whole-cell currents (reflecting the total number of synaptic plus extrasynaptic AMPA receptors) (Fig. 2C). To probe for possible changes in NMDAR-mediated excitatory synaptic transmission, we compared the relative contribution of NMDA versus AMPA receptors to evoked EPSCs. In agreement with the upregulation of NMDAR subunits in hippocampi of *ProSAP1/Shank2*$^{-/-}$ mice (see Fig. 1C, D), we found an approximately 30% increased NMDA/AMPA ratio in mutants versus wild types (Fig. 2D). We further analysed synaptic plasticity. NMDAR-dependent long-term potentiation induced by high-frequency stimulation of the Schaffer collaterals was slightly enhanced in *ProSAP1/Shank2*$^{-/-}$ mice (Fig. 2E). We found no evidence for alterations in long-term depression between genotypes (Supplementary Fig. 9e). As imbalanced excitation/

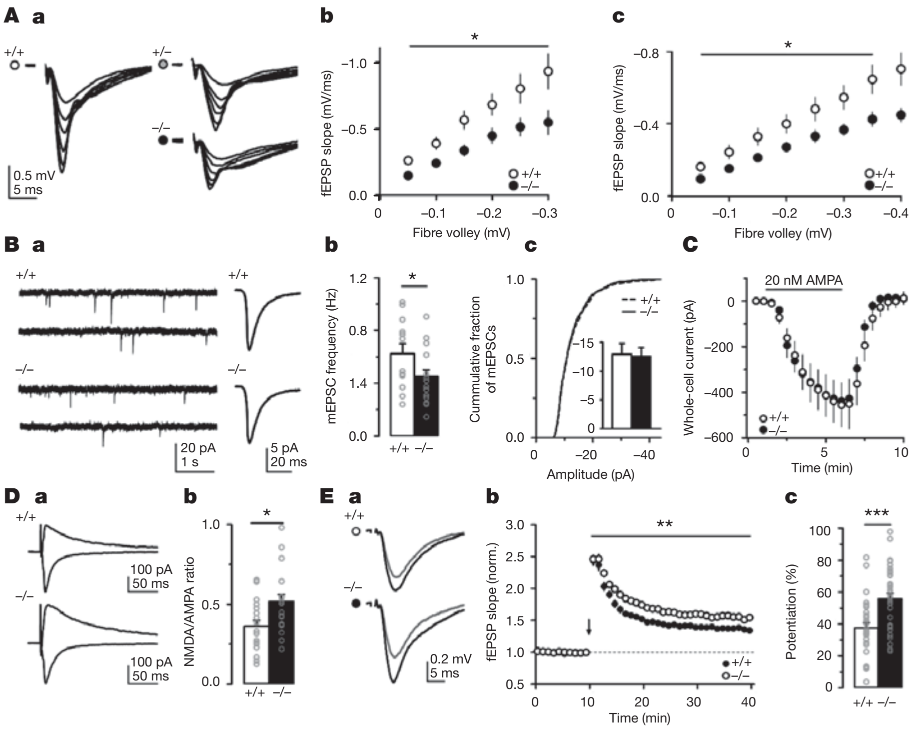


Figure 2 | Imbalanced hippocampal glutamatergic synaptic transmission in *ProSAP1/Shank2*$^{-/-}$ mice. **A**, Input–output curves for basal synaptic transmission. As illustrated in the sample traces (**A**, **a**, averages of six fEPSPs) and in the quantification (**A**, **b**), *ProSAP1/Shank2*$^{-/-}$ (−/−) mice suffer from reduced synaptic transmission compared with wild-type controls (+/+) at the age of P21–P28 (two-way analysis of variance (ANOVA): $P < 0.05$; +/+: number of experiments (n) and number of animals (N) = 8 (3); −/−: n = 11 (4)). This defect is also found in mice that are 3 months of age (**A**, **c**) (two-way ANOVA: $P < 0.05$; +/+: n (N) = 7 (3); −/−: n (N) = 7 (3)). **B**, mEPSCs in CA1 pyramidal cells. **B**, **a**, Sample traces of individual recordings (left) and an average of all mEPSC events (right). **B**, **b**, The frequency of mEPSCs is reduced in *ProSAP1/Shank2*$^{-/-}$ (Student's t-test: $P < 0.05$; +/+: n (N) = 12 (4); −/−: n (N) = 16 (5)). **B**, **c**, Cumulative fraction distribution of mEPSC amplitudes (two-sample Kolmogorov–Smirnov test: $P = 0.96$; +/+: 504 events, n (N) = 12 (4); −/−: 630 events, n = 15 (5)). Inset: mean mEPSC amplitudes (Student's t-test: $P = 0.37$; sample sizes as above). **C**, Whole-cell currents evoked by bath application of 20 nM AMPA (Student's t-test: $P = 0.9$; +/+: n (N) = 8 (3); −/−: n (N) = 9 (3)). **D**, NMDA/AMPA ratios estimated from compound EPSCs evoked at +40 and −60 mV, respectively. As illustrated in the sample traces (**D**, **a**) and the quantification (**D**, **b**), the ratio of synaptic NMDA versus AMPA receptors is significantly increased in *ProSAP1/Shank2*$^{-/-}$ animals (Student's t-test: $P < 0.05$; +/+: n (N) = 18 (8); −/−: n (N) = 19 (6)). **E**, Long-term potentiation is increased in *ProSAP1/Shank2*$^{-/-}$ mice, as evident from the sample traces (**E**, **a**), the average time plot (**E**, **b**) (two-way ANOVA: $P < 0.01$; +/+: n (N) = 30(5); −/−: n = 34 (6)) and the ratio of fEPSP slopes 30 min after versus before induction of long-term potentiation (**E**, **c**) (Student's t-test: $P < 0.001$; sample sizes as above). All data are presented as mean ± s.e.m. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

inhibition ratios have been repeatedly implicated in models of autism[24], we also analysed GABAergic (γ-aminobutyric acid-mediated) synaptic transmission. Frequency and amplitude of inhibitory postsynaptic currents (both miniature and spontaneous) were largely unchanged in *ProSAP1/Shank2*$^{-/-}$ mice (Supplementary Fig. 10a, b). Based on the electrophysiological analyses, we conclude that merely glutamatergic transmission is impaired in *ProSAP1/Shank2* mutants.

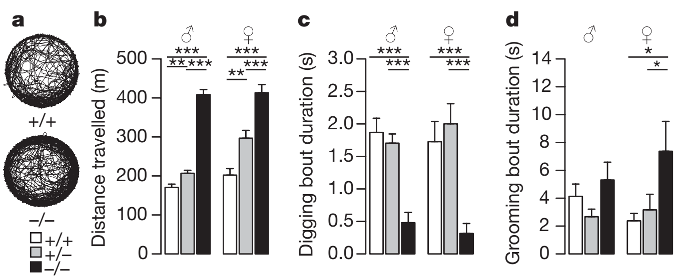


Figure 3 | Increased locomotor activity and stereotypical behaviours in *ProSAP1/Shank2*$^{-/-}$ mice. **a**, Examples of trajectories of a wild-type mouse and a *ProSAP1/Shank2*$^{-/-}$ mouse in 30 min exploration of the open field. **b**, Distance travelled by male and female mice during 30 min free exploration of a circular maze. **c**, Mean digging bout duration in male and female mice. **d**, Mean self-grooming bout duration in male and female mice. Data are presented as mean ± s.e.m. (Mann–Whitney U-tests: *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Unless otherwise specified, ($n_{+/+} = 16$, $n_{+/-} = 16$, $n_{-/-} = 16$) males and ($n_{+/+} = 16$, $n_{+/-} = 16$, $n_{-/-} = 13$) females were tested.

Despite these synaptic abnormalities, *ProSAP1/Shank2*$^{-/-}$ mice displayed functional working memory, motor coordination, olfaction and object recognition (Supplementary Fig. 11a–g). The most remarkable behavioural phenotype was hyperactivity. When compared with wild-type littermates, male and female *ProSAP1/Shank2*$^{-/-}$ mice displayed twice the level of locomotor activity in the open field (Fig. 3a, b) and in other tests (Supplementary Fig. 12a–e). *ProSAP1/Shank2*$^{-/-}$ males displayed an increased level of anxiety during the light–dark box test (Supplementary Fig. 12f, g). Compared with wild types, digging bouts of *ProSAP1/Shank2*$^{-/-}$ mice were significantly shorter (Fig. 3c), and self-grooming in *ProSAP1/Shank2*$^{-/-}$ females was significantly extended (Fig. 3d). These stereotyped behaviours, however, were less severe compared with other mouse models of ASD such as *ProSAP2/Shank3αβ*$^{-/-}$ mutants[17,20] or BTBR T+tf/J mice[25].

We next examined social behaviour. During free same-sex social interactions (resident–intruder), the latency for the first contact did not differ significantly (Fig. 4a), but both male and female *ProSAP1/Shank2*$^{-/-}$ mice had difficulties maintaining social contacts or were less interested in them (Fig. 4b). During free interactions of a tested male mouse with an oestrus C57BL/6 female mouse, the latency for the first contact was significantly longer for *ProSAP1/Shank2*$^{-/-}$ males than wild-type males (Fig. 4a), but no impairment in contact maintenance was detected (Fig. 4b). During the three-chamber test, both male and female mutants displayed a reduction in conspecific recognition or in their interest for social novelty compared with wild types (Supplementary Fig. 13a–d).

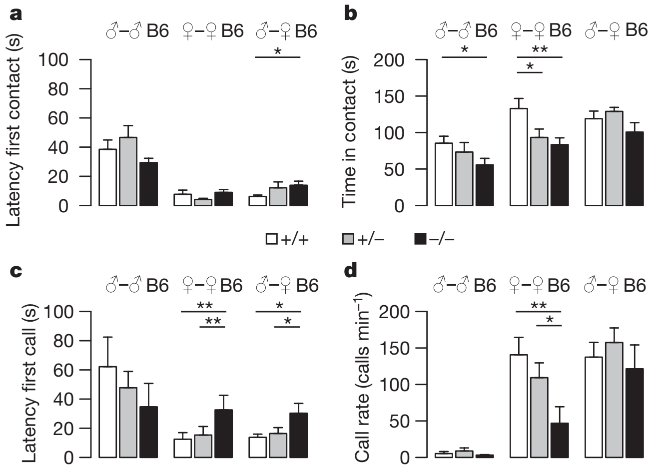


Figure 4 | Abnormalities in social and vocal behaviour of *ProSAP1/Shank2*$^{-/-}$ mice in the resident–intruder test and during the interaction of a male with an oestrus female. **a**, Latency for the first contact in same-sex free interactions (C57BL/6 resident–intruder) when the resident mouse was isolated from weaning on (males) or for 3 days before the experiment (females), and in the interaction of a male with a C57BL/6 oestrus female. **b**, Time spent in contact during same-sex free interactions when the resident mouse was isolated from weaning on (males) or for 3 days before the experiment (females), and in the interaction of a male with an oestrus female ($n_{+/+} = 15$, $n_{+/-} = 16$, $n_{-/-} = 16$). **c**, Latency for the first ultrasonic vocalization in same-sex free interactions when the resident mouse was isolated from weaning on (males) or for 3 days before the experiment (females), and in the interaction of a male with an oestrus female. **d**, Rate of calling during same-sex free interactions when the resident mouse was isolated from weaning on (males) or for 3 days before the experiment (females), and in the interaction of a male with an oestrus female. Data are presented as mean ± s.e.m. (Mann–Whitney *U*-tests: $*P < 0.05$; $**P < 0.01$; $***P < 0.001$). Unless otherwise specified, ($n_{+/+} = 16$, $n_{+/-} = 16$, $n_{-/-} = 16$) males and ($n_{+/+} = 16$, $n_{+/-} = 16$, $n_{-/-} = 13$) females were tested.

During free social interactions or when isolated as pups, mice emit ultrasonic vocalizations. In pups, *ProSAP1/Shank2*$^{-/-}$ females, but not males, called at a significantly higher rate than wild-type females at P4 and P10 (Supplementary Fig. 14a, b). In adults, during male–male social interactions, few calls were recorded and there was no significant difference between genotypes (Fig. 4c, d). In contrast, during female–female interactions, we observed a significantly longer latency to emit the first vocalization (Fig. 4c) and significantly fewer vocalizations (Fig. 4d) for pairs involving a *ProSAP1/Shank2*$^{-/-}$ mouse compared with pairs involving wild-type females. Notably, in pairs involving *ProSAP1/Shank2*$^{-/-}$ mutants, mice uttered more short and unstructured calls and fewer mixed calls than pairs involving wild types (Supplementary Fig. 14c, e). In the socio-sexual context of a male in the presence of an oestrus female, the latency for the first ultrasonic vocalization was significantly longer in pairs involving *ProSAP1/Shank2*$^{-/-}$ males than pairs involving wild types (Fig. 4c). Similarly to females, more short and unstructured calls were emitted (Supplementary Fig. 14d, e).

In conclusion, based on this study and on previous reports, mice lacking any member of the ProSAP/Shank family display recurrent features observed in animal models for ASD: that is, preserved working memory, but increased anxiety and abnormalities in both social interactions and vocalizations (Supplementary Fig. 15)[16–19,26–29].

In summary, here we demonstrate that altered glutamatergic neurotransmission can lead to the core symptoms of ASD. In addition, this study shows that ProSAP1/Shank2 and ProSAP2/Shank3 seem to serve different interrelated functions at excitatory synapses, especially in glutamate receptor targeting/assembly. However, the exact molecular mechanisms are still to be deciphered. In any case, our comparative analysis of mice lacking either ProSAP1/Shank2 or major isoforms of ProSAP2/Shank3 reveals that mutations of very similar proteins within the same synaptic pathway can have different molecular consequences (for example, excess or deficit of glutamate receptors), but both lead to abnormal social and vocal behaviours. Future studies should tell whether gene- or pathway-specific therapies are necessary to modulate or even reverse the pathophysiology of ASD.

METHODS SUMMARY

Biochemistry, Golgi staining and electron microscopy. A subfractionation protocol[30] was performed to obtain subcellular fractions from brain tissue of juvenile and/or adult wild-type, *ProSAP1/Shank2* and *ProSAP2/Shank3* mutant mice from both sexes. After immersion in Golgi–Cox solution for 21 days, adult brains were cut in 200 µm sagittal sections to develop Golgi–Cox staining for analysis of spine density. Further, adult mice were perfused, brains were dissected out, stained and cut in ultrathin sections to be examined by electron microscopy.

Electrophysiology. Extracellular field and whole-cell patch-clamp recordings were performed in horizontal hippocampal slices from mice of both sexes. Evoked postsynaptic responses were induced by electrical stimulation of Schaffer collaterals in CA1 stratum radiatum. fEPSPs were recorded in stratum radiatum. Long-term potentiation was induced by a single tetanus of 100 pulses at 100 Hz. Long-term depression was induced by 15 min paired pulse stimulation at 1 Hz with 50 ms between single pulses. mEPSCs, whole-cell AMPA currents and inhibitory postsynaptic currents (IPSCs) were recorded in whole-cell patch-clamp configuration from CA1 pyramidal cells voltage-clamped at −60 mV. For estimation of NMDA/AMPA ratios, compound EPSCs were evoked at −60 and +40 mV.

Behavioural analysis. Three cohorts of mice (C57BL/6 background) were tested. Cohort 1 included pups for the developmental study to examine pup vocal behaviour, motor coordination, olfaction and developmental milestones. Adult behaviour was tested on cohort 2 in the following order: light–dark anxiety test, open field, Y-maze, three-chamber test, self-directed and digging behaviours, resident–intruder test, male behaviour in presence of an oestrus female, buried-food finding test and object recognition. Cohort 3 was used for the general neurological examination, juvenile body weight and to analyse motor coordination.

All animal procedures were in accordance with institutional, state and government regulations (Tübingen: O.103; Berlin: LAGeSo, T0100/03; Paris: CEEA Ile-de-France Comité 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 November 2011; accepted 8 March 2012.

Published online 29 April; corrected 13 June 2012 (see full-text HTML version for details).

1. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9,** 341–355 (2008).
2. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34,** 27–29 (2003).
3. Etherton, M. R., Tabuchi, K., Sharma, M., Ko, J. & Südhof, T. C. An autism-associated point mutation in the neuroligin cytoplasmic tail selectively impairs AMPA receptor-mediated synaptic transmission in hippocampus. *EMBO J.* **30,** 2908–2919 (2011).
4. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
5. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).
6. Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81,** 1289–1297 (2007).
7. Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **150B,** 421–424 (2009).
8. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39,** 25–27 (2007).
9. Berkel, S. *et al.* Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genet.* **42,** 489–491 (2010).
10. Leblond, C. S. *et al.* Genetic and functional analyses of *SHANK2* mutations provide evidence for a multiple hit model of autism spectrum disorders. *PLoS Genet.* **8,** e1002521 (2012).
11. Baron, M. K. *et al.* An architectural framework that may lie at the core of the postsynaptic density. *Science* **311,** 531–535 (2006).
12. Grabrucker, A. M. *et al.* Concerted action of zinc and ProSAP/Shank in synaptogenesis and synapse maturation. *EMBO J.* **30,** 569–581 (2011).
13. Toro, R. *et al.* Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. *Trends Genet.* **26,** 363–372 (2010).
14. Grabrucker, A. M., Schmeisser, M. J., Schoen, M. & Boeckers, T. M. Postsynaptic ProSAP/Shank scaffolds in the cross-hair of synaptopathies. *Trends Cell Biol.* **21,** 594–603 (2011).
15. Hamdan, F. F. *et al.* Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88,** 306–316 (2011).

16. Bozdagi, O. *et al.* Haploinsufficiency of the autism-associated Shank3 gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism* **1,** 15 (2010).
17. Peca, J. *et al.* Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472,** 437–442 (2011).
18. Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of Shank3. *Hum. Mol. Genet.* **20,** 3093–3108 (2011).
19. Bangash, M. A. *et al.* Enhanced polyubiquitination of Shank3 and NMDA receptor in a mouse model of autism. *Cell* **145,** 758–772 (2011).
20. Chao, H. T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468,** 263–269 (2010).
21. Gemelli, T. *et al.* Postnatal loss of methyl-CpG binding protein 2 in the forebrain is sufficient to mediate behavioral aspects of Rett syndrome in mice. *Biol. Psychiatry* **59,** 468–476 (2006).
22. Boeckers, T. M. *et al.* Proline-rich synapse-associated protein-1/cortactin binding protein 1 (ProSAP1/CortBP1) is a PDZ-domain protein highly enriched in the postsynaptic density. *J. Neurosci.* **19,** 6506–6518 (1999).
23. Berkel, S. *et al.* Inherited and de novo SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. *Hum. Mol. Genet.* **21,** 344–357 (2012).
24. Tabuchi, K. *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318,** 71–76 (2007).
25. Pearson, B. L. *et al.* Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism. *Genes Brain Behav.* **10,** 228–235 (2011).
26. Hung, A. Y. *et al.* Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. *J. Neurosci.* **28,** 1697–1708 (2008).
27. Wohr, M., Roullet, F. I., Hung, A. Y., Sheng, M. & Crawley, J. N. Communication impairments in mice lacking Shank1: reduced levels of ultrasonic vocalizations and scent marking behavior. *PLoS ONE* **6,** e20631 (2011).
28. Silverman, J. L. *et al.* Sociability and motor functions in Shank1 mutant mice. *Brain Res.* **1380,** 120–137 (2011).
29. Ey, E., Leblond, C. S. & Bourgeron, T. Behavioral profiles of mouse models for autism spectrum disorders. *Autism Res.* **4,** 5–16 (2011).
30. Schmeisser, M. J., Grabrucker, A. M., Bockmann, J. & Boeckers, T. M. Synaptic cross-talk between N-methyl-D-aspartate receptors and LAPSER1-beta-catenin at excitatory synapses. *J. Biol. Chem.* **284,** 29146–29157 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Manz, R. Zienecker, S. Gerlach-Arbeiter, N. Damm, H. Riederer, C. Jean, S. Rieckmann, S. Hochmuth and K. Sowa for technical assistance. M.J.S., A.-L.J. and P.T.U. are members of the International Graduate School in Molecular Medicine at Ulm University. M.J.S. is further supported by Baustein 3.2 (L.SBN.0081), E.E. by the Fondation de France and the Agence Nationale de la Recherche (ANR) FLEXNEURIM (ANR09BLAN034003), S.W. and A.V.S. by the Deutsche Forschungsgemeinschaft (DFG) (GRK 1123), A.M.G. by Baustein 3.2 (L.SBN.0083), S.A.S by the DFG (EXC 257), D.S. by the DFG (SFB 618, SFB 665, EXC 257), the Bundesministerium für Bildung und Forschung (BMBF) (BCCN, BFNL) and the Einstein Foundation, M.R.K. by the DFG (SFB 779), C.S.L., R.T., N.T., A.LS. and T.B. by the ANR (ANR-08-MNPS-037-01 - SynGen), Neuron-ERANET (EUHF-AUTISM), Fondation Orange and the Fondation FondaMentale, P.F. by the Bettencourt-Schueller Fondation, R.T., T.B., P.F. by the CNRS Neuroinformatic, E.D.G. by the DFG (SFB 779) and the BMBF (EraNET Neuron), and T.M.B. by the DFG (Bo 1718/3-1 and 1718/4-1; SFB 497/B8).

Author Contributions M.J.S., E.E., J.B., C.S., D.B., S.t.D., K.H.S., D.M., D.S., M.R.K., T.B., E.D.G. and T.M.B. designed the outline of this study. J.B. and B.V.S. generated, and J.B., C.S. and S.A.S. supervised breeding of, the *ProSAP1/Shank2*-mutant mice. J.B. supervised breeding of the *ProSAP2/Shank3*-mutant mice. M.J.S., A.K., A-L.J., P.T.U. and A.M.G. performed all the biochemistry, real-time PCR, Golgi stainings, electron microscopy, transfection of primary neurons and immunohistochemistry, E.E., C.S., D.M., C.S.L., P.F., N.T. and A.LS. the behavioural experiments, and S.W., A.V.S. and D.B. the electrophysiological experiments. E.S. conducted the survival analysis. M.J.S., E.E., S.W., A.V.S., C.S., D.B., D.M., R.T. and A.M.G. performed all data analyses and jointly drafted the manuscript with S.A.S., D.S., M.R.K., T.B., E.D.G. and T.M.B. All authors read and approved the final version. M.J.S., E.E. and S.W. contributed equally to this study. We thank H.-J. Kreienkamp, Hamburg, for providing the pan-Shank antibody '189.3'.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to T.M.B. (tobias.boeckers@uni-ulm.de).

METHODS

Animals. Two mouse C57BL/6 genomic XbaI DNA fragments coding for exon VI and VII of the ProSAP1/Shank2 PDZ domain were subcloned from a bacterial artificial chromosome (pBelo-BAC II 21259) yielding a 16.3-kilobase (kb) chromosomal DNA fragment. For target vector construction, the long arm, coding for exon VI, was cloned as an EcoRI–XbaI fragment (7.7 kb) into pBluescript II SK vector (Thermo Scientific) followed by an XbaI–XhoI fragment (1.1 kb) coding for exon VII in which a loxP sequence was integrated into the unique SphI site. The frt PGK neo frt loxP cassette was cloned into the XhoI site followed by the short arm (2.8 kb) as an XhoI–SphI fragment. The linearized targeting construct was electroporated into recombinant inbred embryonic stem cells (RI-ES cells) (passage 15 (129/SV × 129/SV-CP)) at 25 μF and 400 V (Gene Pulser; Bio-Rad). After electroporation, cells were plated onto culture dishes (100 mm diameter) containing a gamma-irradiated monolayer of G418-resistant PMEF cells. Thirty-two hours later, 350 μg of G418 (Invitrogen) per millilitre and 0.2 μM 2′-deoxy-2′-fluoro-β-D-arabinofuranosyl-5-iodouracil (Moravek Biochemicals and Radiochemicals) were added to the culture medium. The medium was replaced every day, and colonies were picked and analysed 8 days after plating. Cells were expanded in HEPES-buffered Dulbecco's modified Eagle's medium supplemented with 15% fetal bovine serum (Thermo Scientific), 1,000 U of recombinant leukaemia inhibitory factor (Millipore) per millilitre, non-essential amino acids, L-glutamine, β-mercaptoethanol and antibiotics (penicillin $100\,\mathrm{U\,ml^{-1}}$, and streptomycin $100\,\mu\mathrm{g\,ml^{-1}}$). For electroporation, 2×10^7 cells were re-suspended in 20 mM HEPES (pH 7.4), 173 mM NaCl, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM dextrose and 0.1 mM β-mercaptoethanol. Homologous recombination was tested by Southern blot analysis. Embryonic stem cells were transiently transfected with a FRT-recombinase plasmid to delete the selection cassette, leaving a single frt site. Correctly targeted embryonic stem cells were microinjected into 3.5-day-old B6D2F1 blastocysts and transferred to the uteri of 2.5 day pseudopregnant CD1 females. Pregnant mice carried pups to term and born chimaeras were identified by agouti coat colour contribution. For the germ-line transmission, male chimaeras were crossed to C57BL/6 female mice. Heterozygous offsprings (*ProSAP1/Shank2*$^{+/frt}$) were confirmed by Southern blot analysis and further tested by PCR for the presence of the targeted allele. Cre-mediated excision was performed *in vivo* by cross-breeding mice harbouring a CMV promoter driven Cre transgene resulting in heterozygous ProSAP1/Shank2-deficient mice (*ProSAP1/Shank*$^{+/-}$). *ProSAP2/Shank3* mutants were generated by Genoway (Lyon, France) and raised on a C57BL/6 background. The targeting strategy is shown in Supplementary Fig. 6. All mice were kept in specific pathogen-free animal facilities and all mouse procedures were performed in compliance with the guidelines for the welfare of experimental animals issued by the Federal Government of Germany and further approved by the ethical committee of Ile-de-France (CEEA Ile-de-France Comité 1).

Primary antibodies. Two polyclonal ProSAP1/Shank2 antibodies were used for this study. The first one (PRC pab SA6045) was directed against amino acids 355–509, the second one (ppI-SAM pab SA5193) against amino acids 826–1259 of rat ProSAP1/Shank2. Both antisera were produced in rabbits and guinea pigs and have previously been characterized[31]. Furthermore, a novel polyclonal ProSAP2/Shank3 antibody was produced (PRC pab). This antibody was directed against amino acids 781–1009 and 1260–1392 within the proline-rich clusters (PRC) of rat ProSAP2/Shank3. For antibody production, partial complementary DNAs (cDNAs) of rat ProSAP2/Shank3 (base pairs 2116–2800 and 3550–3948, respectively) were each cloned into the bacterial expression vector pGEX-4T (GE Healthcare). The corresponding glutathione *S*-transferase (GST)–ProSAP2/Shank3 fusion proteins were expressed in *Escherichia coli* BL21, purified and injected at the same time in rabbits to generate antiserum. For all experiments conducted in this study, both ProSAP1/Shank2 antisera and the ProSAP2/Shank3 antiserum were each purified against the other family members. For this purpose, COS-7 cells were transfected with pEGFP–Shank1, pEGFP–ProSAP1/Shank2 and/or pEGFP–ProSAP2/Shank3 constructs (all constructs have been described previously[32]). The corresponding fusion proteins were collected, coupled to magnetic beads and loaded onto a column-based system (Miltenyi Biotec). The antisera were then purified by having them flow subsequently through the appropriate columns (the green fluorescent protein (GFP)–Shank1 and GFP–ProSAP2/Shank3 columns in case of the ProSAP1/Shank2 antisera, the GFP–Shank1 and GFP–ProSAP1/Shank2 columns in case of the ProSAP2/Shank3 antiserum). Two different pan-Shank antibodies were further used in this study. The first, pan-Shank '189.3' was directed against the PDZ domain of human Shank1 and is known to detect Shank1 as well as ProSAP1/Shank2 (ref. 33). The second, pan-Shank 'Clone N23B/49', was purchased from Millipore and directed against the SH3/PDZ domains of rat ProSAP1/Shank2 (amino acids 84–309). It is known to recognize all three ProSAP/Shank proteins, though. (The antigen sequence is approximately 75% identical to the corresponding sequence in ProSAP2/Shank3 and approximately 70% to that in Shank1.) The following antibodies were purchased from commercial suppliers: PSD95 (Abcam), β3-tubulin (Covance), GluN2A (NR2A), GluN2B (NR2B) (both Millipore), GKAP/SAPAP, Shank1 (IF) (Novus Biologicals), β-actin, GluN1 (NR1), Shank1 (WB) (all from Sigma-Aldrich), pan-GluA, GluA1, GluA2, GluA3 (all from Synaptic Systems).

Biochemistry and quantitative immunoblot analyses. To obtain the subcellular fractions from mouse brain analysed in this study (homogenates, soluble fractions, crude synaptosomal fractions, synaptic plasma membranes, one-triton extracted PSD fractions), a subcellular fractionation procedure was performed as described previously[30] with minor modifications. In brief, tissue from mouse brain of both sexes (either brain regions and/or whole brain) was homogenized in HEPES-buffered sucrose (320 mM sucrose, 5 mM HEPES, pH 7.4) containing protease inhibitor mixture (Roche). The homogenate was taken for analysis and/or further centrifuged at 1,000*g* for 10 min at 4 °C to remove cell debris and nuclei. The supernatant was spun for 20 min at 12,000*g* to obtain the soluble and the crude synaptosomal fraction (pellet P2). This fraction was used for the broad analysis of scaffold and receptor protein levels throughout various brain regions at two developmental stages (P25, P70). For some experiments, a pooled P2 fraction from 10 whole mouse brains was further fractionated by sucrose density gradient centrifugation (0.8/1.0/1.2 sucrose) at 200,000*g* for 2 h at 4 °C. Purified synaptosomes/synaptic plasma membranes were collected at the 1.0–1.2 M interface. To obtain the one-triton extracted PSD fraction, synaptosomes were re-suspended in five volumes of 1 mM Tris pH 8.1, stirred for 15 min on ice in buffer containing 0.5% Triton X-100 and centrifuged at 33,000*g* for 30 min. Equal amounts of 10–20 μg protein per lane were separated by SDS–polyacrylamide gel electrophoresis, stained with Coomassie gel staining solution (Thermo Scientific) or blotted onto polyvinylidene fluoride or nitrocellulose membranes using standard protocols. After incubation with specific primary antibodies, immunoreactivity was visualized on X-ray film (GE Healthcare) using HRP-conjugated secondary antibodies (Dako) and a SuperSignal detection system (Thermo Scientific). For quantification, the films were scanned and the grey value of each band was analysed by ImageJ (National Institutes of Health, http://rsb.info.nih.gov/ij/) and normalized to the grey value of β-actin and/or β3-tubulin.

Immunohistochemistry. Immunohistochemical stainings were performed on frozen brain sections (5 μm) of adult mice from both sexes. Ice-cold methanol (−20 °C) was used for fixation. Cell membranes were permeabilized with Triton X-100 (0.5% Triton X-100 in 10 mM PBS) followed by blocking with hydrogen peroxide solution (to inactivate endogenous peroxidase) and 2% BSA (to block unspecific binding sites). The sections were subsequently incubated with the first antibody in 0.5% BSA overnight at 4 °C in the wet chamber. For diaminobenzidene staining, the sections were incubated with a biotinylated secondary antibody and detection was ascertained using the ABC Vectastain Kit (Vector Laboratories) and DAB solution (0.6% DAB, 0.1% hydrogen peroxide in 5 mM Tris/HCl, pH 7.4). For immunofluorescence staining, the sections were incubated with an Alexa Fluor 568 (red) fluorescence labelled secondary antibody (Invitrogen) and cell nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI).

Golgi staining. Dissected adult mouse brains from both sexes were immersed for 21 days in Golgi–Cox solution (1% potassium dichromate, 1% mercuric chloride, 8% potassium chromate). The brains were subsequently dehydrated and sagittal sections (200 μm) were cut using a vibratome. Golgi–Cox staining was developed by incubation in 16% ammonia for 30 min before embedding. Z-stack images were taken using an upright Axioscope (Zeiss). For quantitative analysis of spine density, we analysed at least three neurons (pyramidal neurons from CA1 hippocampus) from at least three independent wild-type and *ProSAP1/Shank2*$^{-/-}$ or wild-type and *ProSAP2/Shank3αβ* $^{-/-}$ littermate pairs.

Electron microscopy. Adult mice from both sexes were transcardially perfused with fixative (2% paraformaldehyde, 2.5% glutaraldehyde, 1% saccharose in 0.1 M cacodylate buffer, pH 7.3) and brains were dissected out and postfixed overnight (2% glutaraldehyde, 1% saccharose in 0.1 M cacodylate buffer). After dehydration and staining with 2% uranyl acetate, the material was embedded in epoxy resin. Ultrathin sections were cut using an ultramicrotome. After lead citrate staining, the sections (CA1 hippocampus) were examined using an EM 10 electron microscope. For quantitative analysis of PSD length and thickness, we analysed at least three independent wild-type and *ProSAP1/Shank2* $^{-/-}$ or wild-type and *ProSAP2/Shank3αβ* $^{-/-}$ littermate pairs.

Real-time quantitative PCR. Isolation of total RNA from mouse brain tissue was performed using the RNeasy kit as described by the manufacturer (Qiagen). Isolated RNA was eluted in 20 μl of RNase-free water and stored at −80 °C. For the reverse transcriptase (RT)-mediated PCR studies, first strand synthesis and real-time quantitative RT–PCR amplification were performed in a one-step, single-tube format using the QuantiFast SYBR Green RT–PCR kit (Qiagen). Thermal cycling and fluorescent detection were performed using the Rotor-Gene-Q real-time PCR machine (model 2-Plex HRM) (Qiagen). The qRT–PCR

was assayed in 0.1 ml strip tubes in a total volume of 20 μl reaction mixture containing 1 μl of undiluted total RNA, 2 μl of QuantiTect Primer Assay oligonucleotides, 10 μl of 2× QuantiFast SYBR Green RT–PCR Master Mix supplemented with ROX (5-carboxy-X-rhodamine) dye, 6.8 μl of RNase-free water (supplied with the kit) and 0.2 μl of QuantiFast RT Mix. RT. Amplification conditions were as follows: 10 min at 50 °C and 5 min at 95 °C, followed by 40 cycles of PCR for 10 s at 95 °C for denaturation, 30 s at 60 °C for annealing and elongation (one-step). During the extension, real-time fluorescence measurements were recorded by the PCR machine, thus monitoring real-time PCR amplification by quantitative analysis of the fluorescence emission. The SYBR Green I reporter dye signal was measured against the internal passive reference dye (ROX) to normalize non-PCR-related fluctuations in fluorescence that occur from reaction tube to reaction tube. Resulting data were analysed using the hydroxymethylbilane synthase gene as an internal standard to normalize transcript levels. Cycle threshold (ct) values were calculated by Rotor-Gene-Q Software (version 2.0.2). Cycle threshold values indicate the PCR cycle number at which the measured fluorescence of the indicator dye (SYBR Green I), accordant to the quantity of amplified PCR products, is increasing in a linear fashion above background. All qRT–PCR reactions were run in duplicates in three independent experiments, and mean ct values for each reaction were taken into account for calculations of data analysis. To ascertain primer specificity, a melting curve was obtained for the amplicon products to determine their melting temperatures. Melting curve was driven from 60 °C to 95 °C rising in 1 °C steps whereas fluorescence was recorded continuously. For negative controls and to check for reagent contamination, a complete reaction mixture was used in which the RNA sample was replaced by RNase-free water. Real-time quantitative PCR was performed using oligonucleotides to investigate expression of *ProSAP1/Shank2* and *ProSAP2/Shank3* (validated primer pairs, QuantiTect primer assay, Qiagen) in tissue from wild-type littermates, *ProSAP1/Shank2*$^{+/-}$, *ProSAP1/Shank2*$^{-/-}$ *and ProSAP2/Shank3*$\alpha\beta^{-/-}$ mutants. All consumables used for the extraction of total RNA and real-time PCR analysis were purchased from Qiagen.

Primary hippocampal cultures from rat. Primary cell culture of rat hippocampal neurons (embryonic day 18 (E18)) was performed as described previously[32]. Hippocampal neurons were seeded on poly-L-lysine (0.1 mg ml^{-1}; Sigma-Aldrich) coated coverslips at a density of 3×10^4 cells per well. Cells were grown in Neurobasal medium, complemented with B27 supplement, 0.5 mM L-glutamine, and 100 U ml^{-1} penicillin/streptomycin (all Invitrogen) and maintained at 37 °C in 5% CO_2. Neurons were transfected with ProSAP1/Shank2 short hairpin RNA (shRNA) (sequence: TG CCT TCA CCA AGA AGG AA) and/or scrambled control constructs[32] at 12 days *in vitro* (DIV12). After fixation at DIV14 and immunostaining with the appropriate antibodies, pictures were taken with an upright Axioscope microscope equipped with a Zeiss CCD camera. Quantification of fluorescence signals was performed using ImageJ (National Institutes of Health, http://rsb.info.nih.gov/ij/). For evaluation, fluorescent puncta (ProSAP2/Shank3, Shank1, GluN1, GluN2A, GluN2B, GluA1, GluA2 and GluA3) along dendrites within the field of view were analysed and the signal intensity values of GFP co-localizing (shRNA-ProSAP1/Shank2 or scrambled RNAi expressing cells) and non-GFP co-localizing puncta compared.

Statistical analysis for the methods mentioned above was performed using Microsoft Excel for Macintosh and data were tested for significance using two-tailed, Student's *t*-test and ANOVA. *P* values < 0.05 were stated as significant (**P* < 0.05, ***P* < 0.01, ****P* < 0.001).

Electrophysiology. *ProSAP1/Shank2* mutants were raised on a C57BL/6 background and wild-type littermates were used as a control in all experiments. The experimenters were blind to the genotype of the tested animals for data collection and analyses. Hippocampal slices were prepared from animals of both sexes aged P21–P28 (unless indicated otherwise), and as previously described[34]. Briefly, mice were anaesthetized with isoflurane and decapitated. Brains were rapidly removed and transferred to ice-cold artificial cerebrospinal fluid (ACSF) slicing solution containing (in mM): 87 NaCl, 26 $NaHCO_3$, 50 sucrose, 25 glucose, 2.5 KCl, 1.25 NaH_2PO_4, 3 $MgCl_2$, 0.5 $CaCl_2$, pH 7.4. Tissue blocks containing the hippocampus were mounted on a Vibratome (Leica VT1200) and cut into horizontal slices of 300 μm. Slices were incubated in slicing solution at 35 °C for 30 min, cooled to room temperature and transferred to ACSF containing (in mM): 119 NaCl, 26 $NaHCO_3$, 10 glucose; 2.5 KCl, 2.5 $CaCl_2$, 1.3 $MgCl_2$, 1 NaH_2PO_4. All ACSF was equilibrated with carbogen (95% O_2, 5% CO_2). Slices were stored under submerged conditions for 30 min to 6 h before being transferred to a submerged recording chamber (Luigs and Neumann) where they were perfused with ACSF maintained at room temperature at a rate of 3–4 ml min^{-1}.

Extracellular field and whole-cell patch-clamp recordings were performed. Stimulation and recording pipettes were pulled from borosilicate glass capillaries (Harvard Apparatus; 1.5 mm outside diameter, with a micropipette electrode puller (DMG Universal Puller)). Evoked postsynaptic responses were induced by stimulating Schaffer collaterals (0.1 Hz) in CA1 stratum radiatum with pipettes of 20 μm tip diameter (filled with ACSF), using a stimulus isolater (ISO-flex, A.M.P.I.). fEPSPs were recorded with the same kind of pipettes that were placed in stratum radiatum. fEPSP slopes were determined as d*V*/d*t* of the 20–80% amplitude from averages of five individual traces. Long-term potentiation was induced by a single tetanus of 100 pulses at 100 Hz. For long-term depression experiments, CA1 was isolated by a microcut set at the border of CA2/CA3 immediately before the experiment, and recordings were made in the presence of 1 μM gabazine. Long-term depression was induced by 15 min paired-pulse stimulation at 1 Hz with 50 ms between single pulses. For whole-cell patch-clamp recordings, pipettes had resistances of 2–3 MΩ. CA1 pyramidal cells were held at −60 mV in voltage-clamp mode (not corrected for liquid junction potential). Series resistance (not compensated) was constantly monitored and was not allowed to increase beyond 22 MΩ or change more than 20% during the experiment. mEPSCs and whole-cell AMPA currents were recorded in the presence of 1 μM tetrodotoxin, 0.1 mM cyclothiazide and 1 μM gabazine with a potassium-based intracellular recording solution containing (in mM): 135 K-gluconate, 10 HEPES, 0.5 EGTA, 20 KCl, 2 MgATP and 5 phosphocreatine. Osmolarity was 300 mOsm; pH was adjusted to 7.2 with KOH. mEPSC events were detected with a threshold-based algorithm and their amplitudes calculated from a 1 ms time window around the peak. Events were aligned by their rise time before averaging. The mEPSC frequency was determined from a 3 min time window. The analysis of mEPSC amplitudes included the first 42 events of each recorded cell (one cell with fewer events was excluded from amplitude analysis). Cellular input resistance was calculated from the steady-state current measured in response to a hyperpolarizing test pulse of 50 ms duration at a holding potential of −60 mV. NMDA/AMPA ratios were recorded in ACSF containing 4 mM $CaCl_2$, 4 mM $MgCl_2$ and 1 μM gabazine with a caesium-based intracellular recording solution containing (in mM): 145 CsCl, 10 HEPES, 0.2 EGTA, 2 $MgCl_2$, 2 NaATP, 0.5 NaGTP and 5 phosphocreatine. Osmolarity was 305 mOsm; pH was adjusted to 7.2 with CsOH. For estimation of NMDA/AMPA ratios, compound EPSCs were evoked at −60 and +40 mV. Stimulation was adjusted to produce a single-peaked short-latency response. Ten consecutive EPSCs for each holding potential were averaged. The AMPA receptor-mediated component of the EPSC was estimated by measuring the peak amplitude of the averaged EPSC at −60 mV. The NMDAR-mediated component was estimated at +40 mV by measuring the amplitude of the averaged EPSC 75 ms after stimulation. IPSCs were recorded in the presence of 10 μM NBQX and 10 μM APV with a potassium-chloride-based intracellular recording solution containing (in mM): 145 KCl, 10 HEPES, 0.1 EGTA, 2 $MgCl_2$, 2 Na_2ATP. Osmolarity was 305 mOsm; pH was adjusted to 7.3 with KOH. For mIPSC recordings, 1 μM tetrodotoxin was added to the bath. The IPSC frequency was determined from a 2 min time window. The analysis of mIPSC (sIPSC) amplitudes include the first 120 (260) events of each recorded cell. All drugs were purchased from Tocris Bioscience.

Recordings were performed with an Axopatch 700A Amplifier (Axon Instruments). Data were acquired using a BNC-2090 adaptor chassis, digitized at 5 kHz (PCI 6035E A/D Board, National Instruments), filtered at 1 kHz and recorded in IGOR Pro 4.0 using custom made plug-ins (WaveMetrics).

Analyses were performed using custom-written procedures in IGOR Pro (WaveMetrics) and MATLAB (The Mathworks). mEPSCs were analysed with NeuroMatic (http://www.neuromatic.thinkrandom.com). Data in graphs and the text are presented as mean ± s.e.m. Statistical comparisons between groups were performed in GraphPad Prism (GraphPad Software) with unpaired two-tailed Student's *t*-tests and two-way repeated-measures ANOVA. A Kolmogorov–Smirnov test was performed in MATLAB. Results were considered significant at *P* < 0.05. Stimulus artefacts were blanked in sample traces. Sample sizes are given as number of experiments (*n*) and number of animals (*N*). Data in bar graphs are presented as mean with standard error; grey circles show individual data points.

Behavioural studies. The behavioural studies included several cohorts of mice. Cohort 1 (backcrossed for 10 generations on C57BL/6) was used for the developmental study of pups. Cohort 2 included 3- to 6-month-old adult mice (backcrossed for 11 generations on C57BL/6). The general neurological examination and the rotarod experiments were conducted with cohort 3 of 6- to 8-month-old adult mice (backcrossed for 11 generations on C57BL/6). We tested the offspring (that is, wild-type, *ProSAP1/Shank2*$^{+/-}$ and *ProSAP1/Shank2*$^{-/-}$ littermates) of *ProSAP1/Shank2*$^{+/-}$ mice crossings. The experimenters were blind of the genotype of the tested animals for data collection and analyses.

General parameters indicative of the health and neurological state were addressed following the neurobehavioural examination described by Whishaw *et al.*[35] and the tests of the primary screen of the SHIRPA protocol except startle response[36]. For the developmental study, males were separated from pregnant females 1 or 2 days before birth. Births were checked each morning and evening. Pups of both sexes (cohort 1) were individually identified with long-lasting subcutaneous tattoos (green tattoo paste, Ketchum Manufacturing) on the paws on

P1. Each pup was isolated from dam and littermates and placed in a small enclosure with soft plastic surface in a soundproof chamber (temperature $T = 24 \pm 1\,^{\circ}\mathrm{C}$). Audio recordings lasted 5 min and were conducted every 2 days (P2, P4, P6, P8, P10 and P12). Recording hardware (UltraSoundGate 416-200, condenser ultrasound microphone Polaroid/CMPA) and software (Avisoft SASLab Pro Recorder) were from Avisoft Bioacoustics (sampling frequency 300 kHz; fast Fourier transform length 1,024 points; 16-bit format). After each audio recording, developmental milestones (weight, opening of the eyes, extroversion of the ears and incisive eruption) were noted. Tests for general sensory perception and motor coordination (righting reflex (P2–P12), home cage odour preference (P7), reaction to a click sound from a pen 2 cm behind the ears (P14)) were conducted at least 30 min after audio recordings.

Righting reflex. The pup was turned on its back on a flat surface. The latency until it reached a normal position (face down) was measured. When the pup did not succeed in righting up, latency was set at the maximum time allowed (120 s).

Preference for home cage odour. A plastic cage (20 cm × 30 cm × 6 cm) was divided in three zones. One side of the cage was covered with bedding from the nest cage; the other side was covered with fresh bedding. A neutral zone (width 2.5 cm) separated these two zones. The 7-day-old pup was placed in the middle of the neutral zone. Time spent by the nose of the pup in each zone was measured within 60 s. This test was repeated for each pup and the mean of the two tests was used in the analyses.

To examine adult behaviour, mice of both sexes aged between 3 and 6 months (cohort 2) were tested in the following order: light–dark anxiety test, open field, Y-maze, three-chamber test, self-directed and digging behaviour, resident–intruder test, male behaviour in presence of an oestrus female, buried-food finding test and object recognition. All animals were weighed at 4 months of age, and their weight was compared with age- and sex-matched commercially available C57BL/6 mice. At the same age, mice were held by the tail 20 cm above a table for approx. 1 min to check for hindlimb clasping. Males were housed individually from weaning onwards, given their high aggressiveness towards each other. Females were housed in groups (except for the resident–intruder test; see hereafter). At least 2 days elapsed between two consecutive tests. Unless otherwise specified, we tested and analysed 16 males per genotype and 16 females per genotype (but only 13 *ProSAP1/-Shank2*$^{-/-}$ females). Settings were cleaned with soap and water, and dried with paper towels between each mouse. When bedding was used, new fresh bedding was used for each mouse. Tests using audio-recordings were conducted in a soundproof chamber.

Light–dark anxiety test. The setting consisted of a white, brightly illuminated compartment (1300 lx) connected to a black, very dark one (3 lx) through a small door. The tested mouse was introduced into the white compartment and allowed to explore the whole apparatus for 5 min. The latency to enter the dark compartment, the time spent in each compartment and the number of transitions between compartments were measured.

Open field. The tested mouse was allowed to explore freely for 30 min a round, open-field arena of 1 m diameter (100 lx in the centre of the arena). Automatic detection of the mouse (custom software of P. Faure) recorded the total distance travelled, and the time spent in the central zone versus time spent at the periphery of the maze.

Y-maze. The tested mouse was allowed to explore freely a Y-shaped labyrinth for 3 min. The number of visits in each arm of the labyrinth and sequences of entrance in each arm were recorded.

Three-chamber test. A Plexiglas cage was divided in three compartments as previously described[37]. Both side compartments contained an empty perforated cup (side compartments: 150 lx; central compartment: 140 lx). First, the tested mouse was allowed to explore freely the whole setting, with all doors open for 10 min (phase 1). After this habituation period, the mouse was restricted in the central compartment, while an unfamiliar C57BL/6 mouse of the same sex (stranger 1) was placed under one of the cups (sides alternated between each mouse). The tested mouse was then allowed to explore the whole apparatus for 10 min (phase 2). After that, it was restricted to the central compartment while another unfamiliar C57BL/6 mouse of the same sex (stranger 2) was placed under the other cup. The tested mouse could then again freely explore the whole apparatus for 10 min (phase 3). In all three phases, time spent in each compartment and number of transitions between compartments were automatically recorded. Times spent in contact with the cup containing the mouse (stranger 1) and the empty cup were manually measured in phase 2. Times spent in contact with the cup containing the unfamiliar mouse (stranger 2) and in contact with the cup containing the familiar mouse (stranger 1) were manually measured in phase 3.

Self-directed and digging behaviours. The tested mouse was placed in a new test cage (Plexiglas, 50 cm × 25 cm × 30 cm; 100 lx; clean sawdust bedding) in a soundproof chamber. After 10 min habituation, its behaviour was video-recorded for 10 min. We recorded the time spent self-grooming and digging in the bedding, as well as the number of self-grooming and digging bouts.

Resident–intruder test. Female mice were housed individually 3 days before the testing day to increase their social motivation. Males were reared in social isolation from weaning onwards (see above). The tested mouse was left 30 min for habituation in the test cage (Plexiglas, 50 cm × 25 cm × 30 cm; 100 lx; clean sawdust bedding; soundproof chamber[38]). After this time, an unfamiliar C57BL/6 mouse of the same sex was introduced. The two animals were allowed to interact freely for 4 min. Social interactions were videotaped continuously (high-resolution Sony XCD-SX90CR video camera). At the same time, ultrasonic vocalizations were recorded with a condenser ultrasound microphone Polaroid/CMPA, the interface UltraSoundGate 416-200 and the software Avisoft SASLab Pro Recorder from Avisoft Bioacoustics (sampling frequency: 300 kHz; fast Fourier transform length: 1024 points; 16-bit format). Behaviours were encoded manually. We recorded the latency for the first contact, the time spent in contact, as well as the latency for the first ultrasonic vocalization, the call rate and the distribution of the different call types.

Male–female interactions. All males had a previous contact of 3 days with a female at least 3 days before the experiment. The tested male was left in a test cage (Plexiglas, 50 cm × 25 cm × 30 cm; 100 lx; clean sawdust bedding; soundproof chamber) to habituate for 10 min. An unfamiliar C57BL/6 female in oestrus (tested though vaginal smears) was then introduced. Ultrasonic vocalizations and social interactions were recorded for 3 min with the same setting as cited above. The latency for the first contact, the time spent in contact, as well as the latency for the first ultrasonic vocalization, the call rate and the distribution of the different call types, were recorded.

Buried-food finding test. Four days before testing, female mice were housed individually, like males. Each day, all mice were given four pieces of cocoa-flavoured crisped rice cereals (Coco Pops, Kellogg's) with restricted access to their usual food (SAFE) but water *ad libitum*. Mice that did not eat their cocoa-flavoured cereals were excluded from the test. Twelve hours before testing, mice were food-deprived, with water *ad libitum*. On the day of the test, six pieces of Coco Pops were buried 1.5 cm under sawdust bedding in a clean cage. The tested mouse was placed in the opposite corner. The latency for the mouse to retrieve the food was measured (maximum time: 3 min). For each animal tested, a new clean cage with fresh bedding was used.

Object recognition. To test *ProSAP1/Shank2* mutants' ability to differentiate objects, we used the set up of the three-chamber test. In the habituation phase, the mouse was allowed to explore the whole setting freely, without any object inside. After that, the mouse was restricted in the central compartment and one identical electric socket was placed in each side compartment. The mouse was again allowed to explore the setting freely with the two identical objects for 10 min. After this second period, while the mouse was restricted in the central compartment, one of the sockets was replaced by a copper joint. Again, the mouse was allowed to explore the complete setting for 10 min. In all three phases, the time spent in each compartment and the number of transitions between compartments was automatically recorded. Time spent in contact with each object was manually measured in each phase.

Rotarod test. Animals of cohort 3 received two training sessions (3 h interval) on a rotarod apparatus (TSE Systems) with increasing speed from 4 to 40 r.p.m. for 5 min. After 4 days, mice were tested at 16, 24, 32 and 40 r.p.m. constant speed. The latency to fall off the rod was measured[39].

Analyses of audio recordings. In the pup developmental study, given that there were no significant differences between the number of ultrasonic vocalizations automatically detected (pulse train detection analyses from Avisoft SASLab Pro; hold time 7 ms) and the number of calls manually detected (data not shown), we used the automatic detection of calls to compare the global call rate between genotypes for each age. For vocalizations recorded in adult animals, we manually detected the calls in the software Avisoft SASLab Pro (Avisoft; 75% overlap; time resolution 0.853 ms; frequency resolution 293 Hz; Hamming window).

For pups and adults, we labelled the calls with the labelling function of Avisoft SASLab Pro. Each call was classified in one call category of 11, adapted from ref. 40 as follows. (1) Short: duration shorter than 5 ms; frequency range ≤ 6.25 kHz. (2) Flat: duration longer than 5 ms and frequency range ≤ 6.25 kHz. (3) Upward: increase in frequency; frequency range greater than 6.25 kHz with only one direction of frequency modulation. (4) Downward: decrease in frequency; frequency range greater than 6.25 kHz with only one direction of frequency modulation. (5) Modulated: frequency modulations in more than one direction; frequency range greater than 6.25 kHz. (6) Complex: addition of one or more frequency component (not harmonic). (7) One frequency jump: inclusion of one jump in frequency without time gap between the two frequency components. (8) Frequency jumps: inclusion of more than one jump in frequency without time gaps between the two consecutive frequency components. (9) Mixed: inclusion of a noisy ('unstructured') part within a pure tone call. (10) Unstructured: no pure tone component identifiable; 'noisy' calls. (11) Others: include all the calls that did not fit in any of the preceding categories (for example, calls combining features of several of the previous call types).

Unless otherwise specified, we used Mann–Whitney *U*-tests to compare the behavioural traits measured between genotypes or within genotypes because of the non-normal distribution of the data and the limited sample size in many cases. We compared the distribution of the different call types between genotypes using χ^2 tests. All analyses were conducted with the computing and statistical software R (R Developmental Core Team 2009) and data are presented as mean ± s.e.m. For exact values of statistical analyses, see Supplementary Tables.

31. Boeckers, T. M. *et al.* Proline-rich synapse-associated protein-1/cortactin binding protein 1 (ProSAP1/CortBP1) is a PDZ-domain protein highly enriched in the postsynaptic density. *J. Neurosci.* **19,** 6506–6518 (1999).
32. Grabrucker, A. M. *et al.* Concerted action of zinc and ProSAP/Shank in synaptogenesis and synapse maturation. *EMBO J.* **30,** 569–581 (2011).
33. Zitzer, H., Hoenck, H. H., Baechner, D., Richter, D. & Kreienkamp, H. J. Somatostatin receptor interacting protein defines a novel family of multidomain proteins present in human and rodent brain. *J. Biol. Chem.* **274,** 32997–33001 (1999).
34. Schmitz, D., Mellor, J., Breustedt, J. & Nicoll, R. A. Presynaptic kainate receptors impart an associative property to hippocampal mossy fiber long-term potentiation. *Nature Neurosci.* **6,** 1058–1063 (2003).
35. Wishaw, I. Q., Haun, F. & Kolb, B. in *Modern Techniques in Neuroscience* (eds Windhorst, U. & Johansson, H.) 1243–1275 (Springer, 1999).
36. Rogers, D. C. *et al.* Behavioral and functional analysis of mouse phenotype: SHIRPA, a proposed protocol for comprehensive phenotype assessment. *Mamm. Genome* **8,** 711–713 (1997).
37. Nadler, J. J. *et al.* Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav.* **3,** 303–314 (2004).
38. Bourgeron, T., Jamain, S. & Granon, S. in *Contemporary Clinical Neuroscience: Transgenic and Knockout Models of Neuropsychiatric Disorders* (eds Fisch, G.S. & Flint, J.) 151–174 (Humana Press, 2006).
39. Montag-Sallaz, M., Schachner, M. & Montag, D. Misguided axonal projections, NCAM180 mRNA upregulation, and altered behavior in mice deficient for the close homolog of L1 (CHL1). *Mol. Cell. Biol.* **22,** 7967–7981 (2002).
40. Scattoni, M. L. Unusual repertoire of vocalizations in adult BTBR T+tf/J mice during three types of social encounters. *Genes Brain Behav.* **10,** 44–56 (2011).

LETTER

doi:10.1038/nature11208

Autistic-like social behaviour in *Shank2*-mutant mice improved by restoring NMDA receptor function

Hyejung Won[1,2]*, Hye-Ryeon Lee[3]*, Heon Yung Gee[4]*, Won Mah[1,2]*, Jae-Ick Kim[3]*, Jiseok Lee[1,2], Seungmin Ha[1,2], Changuk Chung[1,2], Eun Suk Jung[4], Yi Sul Cho[5], Sae-Geun Park[1], Jung-Soo Lee[4], Kyungmin Lee[6], Daesoo Kim[1], Yong Chul Bae[5], Bong-Kiun Kaang[3,7], Min Goo Lee[4] & Eunjoon Kim[1,2,8,9]

Autism spectrum disorder (ASD) is a group of conditions characterized by impaired social interaction and communication, and restricted and repetitive behaviours. ASD is a highly heritable disorder involving various genetic determinants[1]. Shank2 (also known as ProSAP1) is a multi-domain scaffolding protein and signalling adaptor enriched at excitatory neuronal synapses[2–4], and mutations in the human *SHANK2* gene have recently been associated with ASD and intellectual disability[5]. Although ASD-associated genes are being increasingly identified and studied using various approaches, including mouse genetics[6–16], further efforts are required to delineate important causal mechanisms with the potential for therapeutic application. Here we show that *Shank2*-mutant (*Shank2*$^{-/-}$) mice carrying a mutation identical to the ASD-associated microdeletion in the human *SHANK2* gene exhibit ASD-like behaviours including reduced social interaction, reduced social communication by ultrasonic vocalizations, and repetitive jumping. These mice show a marked decrease in NMDA (*N*-methyl-D-aspartate) glutamate receptor (NMDAR) function. Direct stimulation of NMDARs with D-cycloserine, a partial agonist of NMDARs, normalizes NMDAR function and improves social interaction in *Shank2*$^{-/-}$ mice. Furthermore, treatment of *Shank2*$^{-/-}$ mice with a positive allosteric modulator of metabotropic glutamate receptor 5 (mGluR5), which enhances NMDAR function via mGluR5 activation[17], also normalizes NMDAR function and markedly enhances social interaction. These results suggest that reduced NMDAR function may contribute to the development of ASD-like phenotypes in *Shank2*$^{-/-}$ mice, and mGluR modulation of NMDARs offers a potential strategy to treat ASD.

Mutations in the *SHANK2* gene have recently been identified in individuals with ASD and intellectual disability[5,18]. Among these mutations, one *de novo SHANK2* microdeletion found in ASD leads to loss of exons 6 and 7 and a frame shift, with concomitant removal of the PDZ and following domains in SHANK2 proteins. To explore the possibility that this deletion causes ASD in humans, and to study the mechanisms underlying the development of ASD, we generated transgenic mice carrying a mutation identical to the human microdeletion (exons 6 and 7 deletion and a frame shift), which affects both splice variants of *Shank2* (*Shank2a* and *Shank2b*) in mice (Fig. 1a). The deletion was verified by Southern blotting and various PCR methods (Supplementary Fig. 1). Shank2 proteins were undetectable in the brain (Fig. 1b), and there were no compensatory increases in Shank1 or Shank3 (Supplementary Fig. 1). The *Shank2*$^{-/-}$ mice showed normal reproduction and brain structure (Supplementary Fig. 2).

We first examined whether *Shank2*$^{-/-}$ mice displayed autistic-like impairments in social interaction. In a home-cage social interaction assay, *Shank2*$^{-/-}$ mice showed reduced interaction with normal target mice, as compared with wild-type animals (Supplementary Fig. 3). In a three-chamber social interaction assay, wild-type animals preferred to explore the first novel mouse introduced (stranger 1) over an inanimate object relatively more than did *Shank2*$^{-/-}$ mice (Fig. 1c, d and Supplementary Fig. 4). Next, when the object was replaced by another novel mouse (stranger 2), *Shank2*$^{-/-}$ mice preferred to explore stranger 2 over stranger 1, similar to wild-type animals (Fig. 1e), indicative of normal levels of social novelty recognition. Similar results were obtained when we used juvenile *Shank2*$^{-/-}$ mice (Supplementary Fig. 5). *Shank2*$^{-/-}$ mice had normal olfactory function (Supplementary Fig. 6).

Shank2$^{-/-}$ mice showed impaired spatial learning and memory in the Morris water maze, although novel object recognition memory was normal (Supplementary Fig. 7). These results suggest that *Shank2*$^{-/-}$ mice have partially impaired learning and memory, consistent with the idea that the deletion in exons 6 and 7 in humans causes ASD and mild to moderate intellectual disability[5].

Shank2$^{-/-}$ mice showed impairments in social communication by ultrasonic vocalizations (USVs). When allowed to interact with a novel wild-type female mouse, *Shank2*$^{-/-}$ male mice uttered USVs less frequently than did wild-type animals, and took longer to make the first call (Fig. 1f–h). In a pup retrieval assay, *Shank2*$^{-/-}$ female mice retrieved the pups less efficiently than did wild-type mice (Fig. 1i and Supplementary Fig. 8, Supplementary Movies 1 and 2).

Shank2$^{-/-}$ male animals exhibited other autistic-like abnormalities. When kept alone in stranger-free home cages, *Shank2*$^{-/-}$ mice showed enhanced jumping mostly mixed with upright scrabbling, normal grooming, and decreased digging behaviours (Fig. 1j and Supplementary Movies 3–5). *Shank2*$^{-/-}$ mice also displayed impaired nesting behaviour, hyperactivity in assays including the open field test, anxiety-like behaviour in an elevated plus maze, and increased grooming in a novel object recognition arena (Supplementary Figs 4, 6 and 7). *Shank2*$^{-/-}$ female mice showed similar repetitive jumping, hyperactivity in an open field, and anxiety-like behaviour in an elevated plus maze, although not in a light-dark box (Supplementary Fig. 8). These data collectively suggest that *Shank2*$^{-/-}$ mice show ASD-like behaviours. It should be noted that the hyperactivity and anxiety-like behaviours might contribute to the impaired social interaction in *Shank2*$^{-/-}$ mice by limiting target exploration or evoking anxiety-like responses.

We also characterized heterozygous *Shank2* (*Shank2*$^{+/-}$) mice, because human gene mutations are mostly heterozygotic. *Shank2*$^{+/-}$ mice showed hyperactivity, similar to *Shank2*$^{-/-}$ mice (Supplementary Fig. 9). However, they showed no abnormalities in social interaction,

[1]Department of Biological Sciences, KAIST, Daejeon 305-701, Korea. [2]National Creative Research Initiative Center for Synaptogenesis, KAIST, Daejeon 305-701, Korea. [3]National Creative Research Initiative Center for Memory, Department of Biological Sciences, College of Natural Sciences, Seoul National University, Gwanangno 599, Gwanak-gu, Seoul 151-747, Korea. [4]Department of Pharmacology, Brain Korea 21 Project for Medical Sciences, Severance Biomedical Science Institute, Yonsei University College of Medicine, Seoul 120-752, Korea. [5]Department of Oral Anatomy and Neurobiology, School of Dentistry, Kyungpook National University, Daegu 700-412, Korea. [6]Department of Anatomy, School of Medicine, Brain Science & Engineering Institute, Kyungpook National University, Daegu 700-412, Korea. [7]Department of Brain and Cognitive Sciences, Seoul National University, Seoul 151-747, Korea. [8]Graduate School of Nanoscience and Technology (World Class University), KAIST, Daejeon 305-701, Korea. [9]Center for Synaptic Brain Dysfunctions, Institute for Basic Science, Daejeon 305-811, Korea.
*These authors contributed equally to this work.

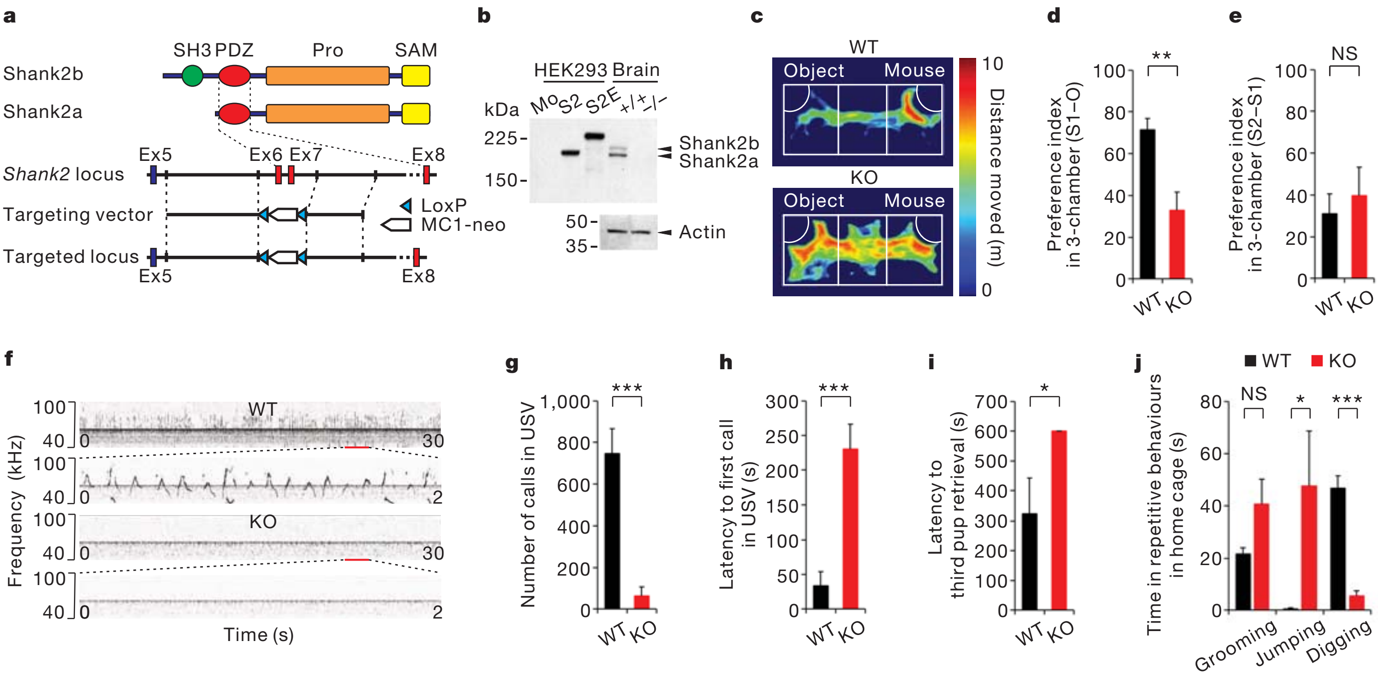


Figure 1 | *Shank2*$^{-/-}$ mice exhibit ASD-like impaired social interaction and social communication, and repetitive jumping. **a**, Targeting of the *Shank2* gene in mice. Ex, exon. **b**, *Shank2*$^{-/-}$ brain lacks expression of both *Shank2a* and *Shank2b* splice variants. Mo, mock; S2, Shank2; S2E, epithelial form of Shank2. **c–e**, Impaired social interaction of *Shank2*$^{-/-}$ mice in three-chamber assays. KO, knockout; WT, wild type. **d**, Social preference (object versus stranger 1 (S1–O)). **e**, Social novelty recognition (stranger 1 versus stranger 2 (S2–S1)). $n = 11$ (WT), 10 (KO). **f–h**, Impaired social communication by USVs in *Shank2*$^{-/-}$ mice. $n = 10$ (WT), 10 (KO). **i**, Impaired pup retrieval in *Shank2*$^{-/-}$ mice. $n = 5$ (WT), 5 (KO). **j**, Stereotypical behaviours in *Shank2*$^{-/-}$ mice. $n = 11$ (WT), 11 (KO). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, NS, not significant. Data represent mean ± standard error.

repetitive behaviours, or anxiety-like behaviours, reflecting intrinsic differences between humans and mice.

Shank2 is an important regulator of excitatory synaptic structure and function[2–4,19]. *Shank2* deletion, however, had minimal effects on excitatory or inhibitory synapses (Supplementary Fig. 10). In addition, electron microscopy revealed that excitatory synapse number and postsynaptic density morphology were unaltered (Supplementary Fig. 11).

We next measured synaptic transmission at hippocampal Schaffer-collateral–CA1-pyramidal (SC–CA1) synapses. Basal excitatory transmission such as input–output and paired-pulse ratio was unchanged in *Shank2*$^{-/-}$ mice (Fig. 2a, b). In addition, spontaneous transmission and membrane excitability were normal in mutant animals (Supplementary Fig. 12). When synaptic plasticity was tested, long-term potentiation (LTP) induced by high-frequency stimulation or theta-burst stimulation was severely impaired in *Shank2*$^{-/-}$ mice (Fig. 2c and Supplementary Fig. 13). Long-term depression (LTD) induced by low-frequency stimulation was completely abolished also in *Shank2*$^{-/-}$ mice (Fig. 2d). Because LTD induced by low-frequency stimulation activates both NMDARs and mGluRs[20], we isolated mGluR LTD by bath-applying (*RS*)-3,5-dihydroxyphenylglycine (DHPG), an agonist of mGluR5, but found no difference between genotypes (Supplementary Fig. 13). This suggests that the observed reductions in LTP and LTD may be due to NMDAR hypofunction.

We thus measured the NMDA/AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) ratio at *Shank2*$^{-/-}$ SC–CA1 synapses. Indeed, the NMDA/AMPA ratio was reduced relative to wild-type synapses (Fig. 2e). Meanwhile, both the decay kinetics of NMDAR excitatory post-synaptic currents (EPSCs) and GluN2B-mediated EPSCs (GluN2B also known as NR2B or Grin2b) were indistinguishable between genotypes, suggesting that GluN2A- (also known as NR2A or Grin2a) and GluN2B-containing NMDARs were equally affected (Supplementary Fig. 14). Given that AMPA receptor (AMPAR)-mediated transmission is normal (Fig. 2a), these results suggest that NMDAR-mediated transmission is selectively decreased. The NMDA/AMPA ratio in the medial prefrontal cortex, however, was unaltered in *Shank2*$^{-/-}$ mice (Supplementary Fig. 14), suggesting that the reduced NMDA/AMPA ratio is not a change uniformly occurring in all brain regions.

Shank2 deletion may also affect NMDAR-associated signalling that critically regulates various synaptic events including LTP and LTD[21,22]. In immunoblot analyses, phosphorylation but not total levels of CaMKII-α/β (T286), ERK1/2 (p42/44) and p38 were significantly reduced in the *Shank2*$^{-/-}$ brain (Supplementary Fig. 15). A similar decrease was observed in phosphorylation of the AMPAR subunit GluA1 (S831 and S845). There were no changes in phosphorylation of PAK1/3 and mTOR, total levels of glutamate receptors (GluN2A, GluA2 (also known as GluR2 or Gria2) and mGluR1/5 (also known as Grm1/5)), or total levels of synaptic scaffolds and signalling adaptors/proteins directly or indirectly associated with Shank2 including PSD-95 (also known as Dlg4), SAP97 (also known as Dlg1), GKAP (also known as Dlgap1), SynGAP1, Homer1, Arhgef6/7 (also known as α/βPIX), GIT1 and PLC-β3. The increase in GluN1 expression may reflect a compensatory increase. These results suggest that Shank2 deficiency leads to impairments in NMDAR-associated signalling.

Reduced NMDAR function and associated signalling may contribute to ASD-like behaviours in *Shank2*$^{-/-}$ mice. To test this hypothesis directly and restore NMDAR function, we used D-cycloserine, a partial agonist at the glycine-binding site of NMDARs, which has been shown to rescue repetitive grooming in neuroligin-1-deficient mice associated with a reduced NMDA/AMPA ratio[23]. We found that D-cycloserine fully recovered the NMDA/AMPA ratio (Fig. 3a). In addition, D-cycloserine-treated *Shank2*$^{-/-}$ mice showed improved social interaction in three-chamber social interaction assays (Fig. 3b–d and Supplementary Fig. 16).

To explore further the association between reduced NMDAR function and ASD-like behaviours in *Shank2*$^{-/-}$ mice, we used 3-cyano-*N*-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide (CDPPB), a membrane-permeable positive allosteric modulator of mGluR5, which increases the responsiveness of mGluR5 to glutamate and enhances

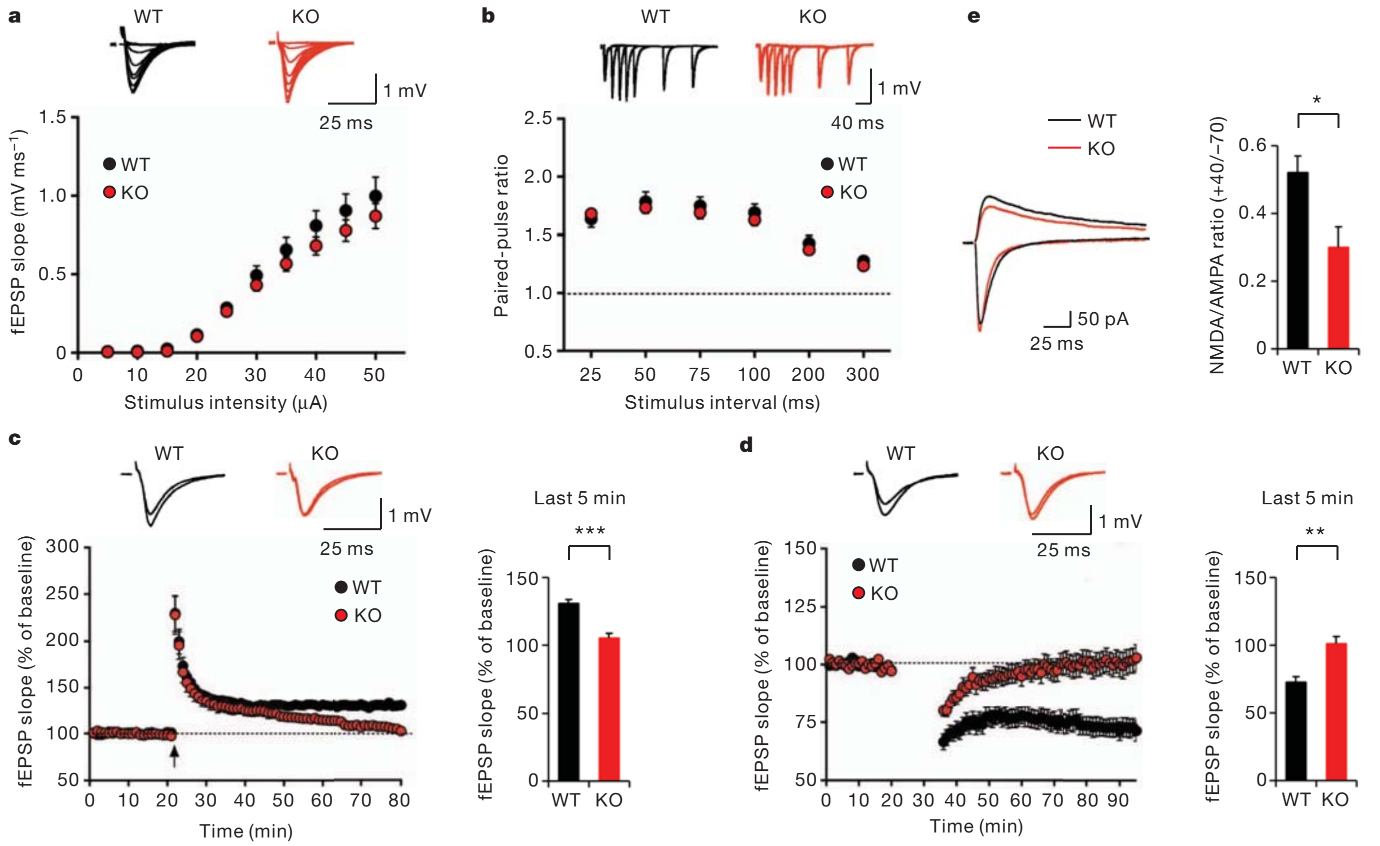


Figure 2 | Impaired NMDAR-dependent synaptic plasticity in *Shank2*$^{-/-}$ mice. **a**, Normal input–output curve at hippocampal SC–CA1 synapses in *Shank2*$^{-/-}$ mice. $n = 7$ (WT), 8 (KO). fEPSP, field excitatory postsynaptic potential. **b**, Normal paired-pulse ratio. $n = 10$ (WT), 8 (KO). **c**, Impaired LTP. $n = 6$ (WT), 8 (KO). **d**, Impaired LTD. $n = 10$ (WT), 9 (KO). **e**, Reduced NMDA/AMPA ratio. $n = 8$ (WT), 8 (KO). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Data represent mean ± standard error.

NMDAR function[17,24,25]. CDPPB has antipsychotic and pro-cognitive activities[17,24,26–28], and facilitates behavioural flexibility[29]. In addition, CDPPB restores reduced excitatory transmission and ERK phosphorylation caused by *Shank3* knockdown[25], and CDPPB and its derivative (VU-29) enhance both LTP and LTD, and spatial learning[27].

Consistent with previous findings, CDPPB also normalized the NMDA/AMPA ratio in *Shank2*$^{-/-}$ brain slices (Fig. 4a). Moreover, CDPPB restored the impaired LTP and LTD at SC–CA1 synapses (Fig. 4b, c), without affecting basal synaptic transmission (Supplementary Fig. 17). Biochemically, CDPPB treatment of *Shank2*$^{-/-}$ mice fully normalized NMDAR signalling in *Shank2*$^{-/-}$ whole brains and also in *Shank2*$^{-/-}$ synaptosomes (Supplementary Figs 18). The smaller extent of signalling deficits in older mice (8 weeks) relative to younger mice (3–4 weeks) may reflect age-dependent reductions in NMDAR-mediated currents and/or compensatory changes in NMDAR signalling.

Behaviourally, *Shank2*$^{-/-}$ mice treated with CDPPB showed substantial recoveries in social interaction to a greater extent than those treated with D-cycloserine, while having no effect on social novelty recognition (Fig. 4d–f and Supplementary Fig. 19). A lower dose of CDPPB did not rescue impaired social interaction (Supplementary Fig. 20), indicative of a dose-dependent action. Notably, CDPPB did not rescue impaired pup retrieval, repeated jumping, anxiety-like behaviours and hyperactivity (Supplementary Fig. 21), suggesting that CDPPB selectively rescues social interaction, but not other behaviours.

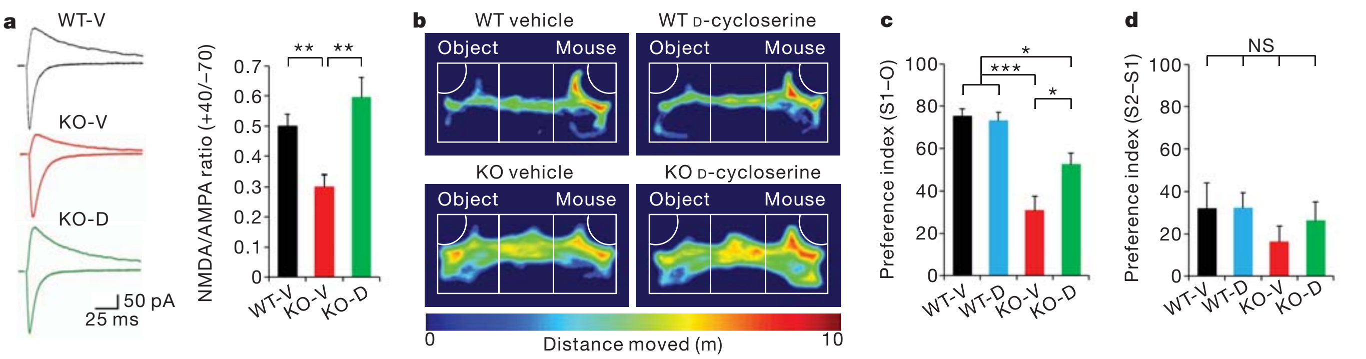


Figure 3 | D-cycloserine normalizes NMDAR function and improves social interaction in *Shank2*$^{-/-}$ mice. **a**, D-cycloserine (20 µM) recovers the reduced NMDA/AMPA ratio. $n = 13$ (wild type, vehicle (WT-V)), 9 (knockout, vehicle (KO-V)), 9 (knockout, D-cycloserine (KO-D)). **b–d**, D-cycloserine (20 mg kg^{-1}) partially normalizes the impaired three-chamber social interaction in *Shank2*$^{-/-}$ mice. $n = 9$ (WT-V), 10 (WT-D), 11 (KO-V), 10 (KO-D). $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, NS, not significant. Data represent mean ± standard error.

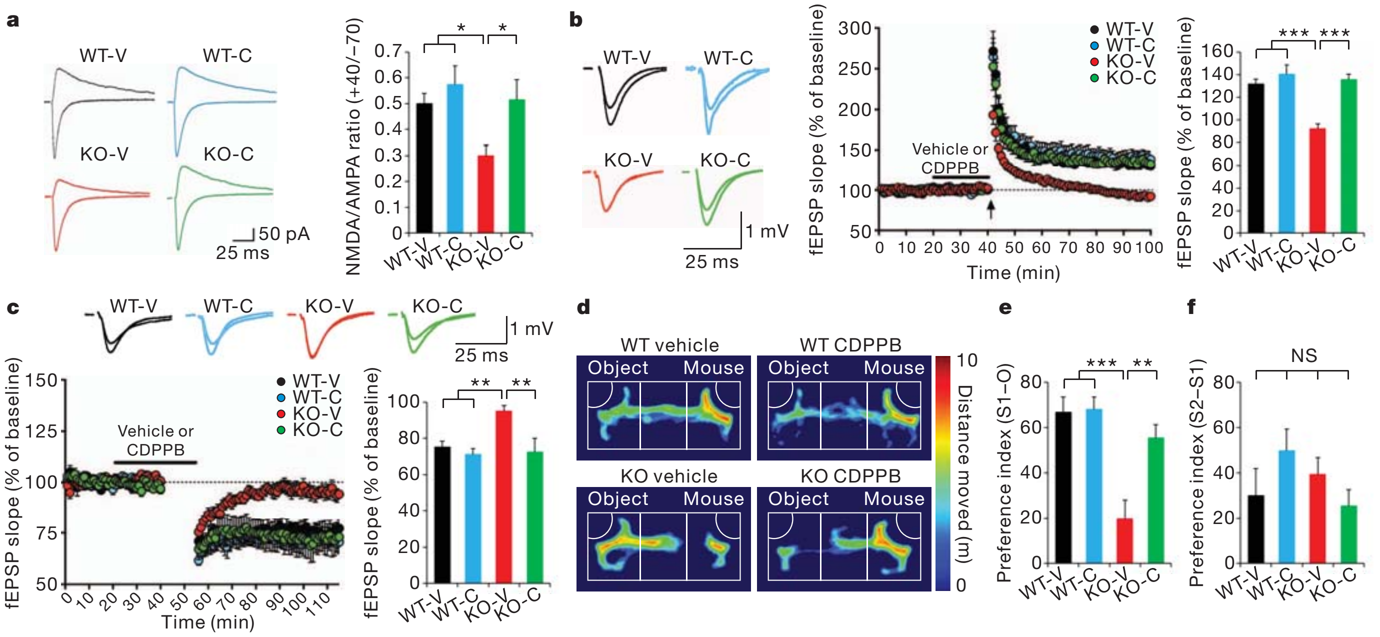


Figure 4 | CDPPB normalizes NMDAR function and substantially improves social interaction in *Shank2*$^{-/-}$ mice. a, CDPPB (10 μM) restores the NMDA/AMPA ratio at *Shank2*$^{-/-}$ SC–CA1 synapses. $n = 13$ (wild type, vehicle (WT-V)), 8 (wild type, CDPPB (WT-C)), 9 (knockout, vehicle (KO-V)), 9 (knockout, CDPPB (KO-C)). **b**, **c**, CDPPB (10 μM) recovers impaired LTP and LTD. $n = 5$ (WT-V), 5 (WT-C), 5 (KO-V), 6 (KO-C) for high-frequency stimulation LTP, and $n = 5$ (WT-V), 6 (WT-C), 7 (KO-V), 6 (KO-C) for low-frequency stimulation LTD. **d**–**f**, *Shank2*$^{-/-}$ mice treated with CDPPB (10 mg kg^{-1}) show substantially improved social interaction in three-chamber assays. **e**, **f**, Quantification of the results in **d**. $n = 8$ (WT-V), 8 (WT-C), 9 (KO-V), 9 (KO-C). $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, NS, not significant. Data represent mean ± standard error.

These results, together with the D-cycloserine results, suggest that reduced NMDAR function and signalling lead to impaired social interaction in *Shank2*$^{-/-}$ mice, although NMDAR-independent mechanisms may also have a role.

Recently, another line of *Shank2*$^{-/-}$ mice produced by deleting exon 7 has been reported to display multiple phenotypes, including reduced spine number, reduced basal transmission, elevated NMDAR currents and ASD-like behavioural changes[30]. Given that our *Shank2*$^{-/-}$ mice lack both exons 6 and 7, the observed differences in mouse phenotype might reflect the differences in genetic deletions and are in line with the different ASD symptoms observed in humans[5]. In addition, the fact that both reduced and enhanced NMDAR functions lead to the same ASD-like phenotypes in mice suggest that maintaining normal levels of NMDAR function is important.

We have demonstrated that NMDAR function is an important mechanism underlying the development and rescue of ASD-like phenotypes in *Shank2*$^{-/-}$ mice, and that mGluR5 may be a novel target for the treatment of ASD involving altered NMDAR function.

METHODS SUMMARY

Animals and statistical analysis. Numbers, genders and ages of mice used for behavioural and other assays are summarized in Supplementary Table 1. All behavioural and electrophysiological assays were performed and analysed in a blind manner. Statistical analyses were performed using SPSS 12.0 (SPSS) and OriginPro (OriginLab), and details of the results are described in Supplementary Table 2.

Experimental details of mouse generation and characterization by behavioural, electrophysiological, biochemical and immunohistochemical analyses are described in Supplementary Methods.

Received 16 November 2011; accepted 11 May 2012.

1. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).
2. Sheng, M. & Kim, E. The shank family of scaffold proteins. *J. Cell Sci.* **113,** 1851–1856 (2000).
3. Boeckers, T. M., Bockmann, J., Kreutz, M. R. & Gundelfinger, E. D. ProSAP/Shank proteins—a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *J. Neurochem.* **81,** 903–910 (2002).
4. Ehlers, M. D. Synapse structure: glutamate receptors connected by the shanks. *Curr. Biol.* **9,** R848–R850 (1999).
5. Berkel, S. *et al.* Mutations in the *SHANK2* synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genet.* **42,** 489–491 (2010).
6. Ehninger, D. & Silva, A. J. Rapamycin for treating tuberous sclerosis and autism spectrum disorders. *Trends Mol. Med.* **17,** 78–87 (2011).
7. Ramocki, M. B. & Zoghbi, H. Y. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* **455,** 912–918 (2008).
8. Jamain, S. *et al.* Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc. Natl Acad. Sci. USA* **105,** 1710–1715 (2008).
9. Tabuchi, K. *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318,** 71–76 (2007).
10. Bozdagi, O. *et al.* Haploinsufficiency of the autism-associated *Shank3* gene leads to deficits in synaptic function, social interaction, and social communication. *Mol. Autism* **1,** 15 (2010).
11. Peça, J. *et al.* *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* **472,** 437–442 (2011).
12. Bangash, M. A. *et al.* Enhanced polyubiquitination of Shank3 and NMDA receptor in a mouse model of autism. *Cell* **145,** 758–772 (2011).
13. Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms of *Shank3*. *Hum. Mol. Genet.* **20,** 3093–3108 (2011).
14. Silverman, J. L. *et al.* Sociability and motor functions in *Shank1* mutant mice. *Brain Res.* **1380,** 120–137 (2011).
15. Südhof, T. C. Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* **455,** 903–911 (2008).
16. Berkel, S. *et al.* Inherited and *de novo* SHANK2 variants associated with autism spectrum disorder impair neuronal morphogenesis and physiology. *Hum. Mol. Genet.* **21,** 344–357 (2012).
17. Gregory, K. J., Dong, E. N., Meiler, J. & Conn, P. J. Allosteric modulation of metabotropic glutamate receptors: structural insights and therapeutic potential. *Neuropharmacology* **60,** 66–81 (2011).
18. Leblond, C. S. *et al.* Genetic and functional analyses of *SHANK2* mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet.* **8,** e1002521 (2012).
19. Hayashi, M. K. *et al.* The postsynaptic density proteins Homer and Shank form a polymeric network structure. *Cell* **137,** 159–171 (2009).
20. Oliet, S. H., Malenka, R. C. & Nicoll, R. A. Two distinct forms of long-term depression coexist in CA1 hippocampal pyramidal cells. *Neuron* **18,** 969–982 (1997).
21. Zhu, J. J., Qin, Y., Zhao, M., Van Aelst, L. & Malinow, R. Ras and Rap control AMPA receptor trafficking during synaptic plasticity. *Cell* **110,** 443–455 (2002).
22. Shepherd, J. D. & Huganir, R. L. The cell biology of synaptic plasticity: AMPA receptor trafficking. *Annu. Rev. Cell Dev. Biol.* **23,** 613–643 (2007).
23. Blundell, J. *et al.* Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J. Neurosci.* **30,** 2115–2129 (2010).
24. Uslaner, J. M. *et al.* Dose-dependent effect of CDPPB, the mGluR5 positive allosteric modulator, on recognition memory is associated with GluR1 and CREB

phosphorylation in the prefrontal cortex and hippocampus. *Neuropharmacology* **57,** 531–538 (2009).
25. Verpelli, C. *et al.* Importance of Shank3 protein in regulating metabotropic glutamate receptor 5 (mGluR5) expression and signaling at synapses. *J. Biol. Chem.* **286,** 34839–34850 (2011).
26. Kinney, G. G. *et al.* A novel selective positive allosteric modulator of metabotropic glutamate receptor subtype 5 has *in vivo* activity and antipsychotic-like effects in rat behavioral models. *J. Pharmacol. Exp. Ther.* **313,** 199–206 (2005).
27. Ayala, J. E. *et al.* mGluR5 positive allosteric modulators facilitate both hippocampal LTP and LTD and enhance spatial learning. *Neuropsychopharmacology* **34,** 2057–2071 (2009).
28. Auerbach, B. D., Osterweil, E. K. & Bear, M. F. Mutations causing syndromic autism define an axis of synaptic pathophysiology. *Nature* **480,** 63–68 (2011).
29. Darrah, J. M., Stefani, M. R. & Moghaddam, B. Interaction of *N*-methyl-D-aspartate and group 5 metabotropic glutamate receptors on behavioral flexibility using a novel operant set-shift paradigm. *Behav. Pharmacol.* **19,** 225–234 (2008).
30. Schmeisser, M. J. *et al.* Autistic-like behaviours and hyperactivity in mice lacking ProSAP1/Shank2. *Nature* http://dx.doi.org/10.1038/nature11015 (29 April 2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank Macrogen for assistance in the production of mice. This work was supported by the National Creative Research Initiative Program, WCU program (R31-2008-000-10071-0), and Institute for Basic Science (to E.K.), the National Research Foundation of Korea (to M.G.L.; grant 2012-0000812), the National Creative Research Initiative Program & WCU program (to B.-K.K.; 2007-0054846), the Basic Science Research Program through the National Research Foundation of Korea (to K.L. and Y.C.B.; 2011-0028240), and the National Leading Research Laboratory Program (to D.K.; 2011-0028772). H.-R.L. and J.-I.K. are supported by the BK21 fellowship, and H.W. is supported by the TJ Park Doctoral Fellowship and National Junior Research Fellowship.

Author Contributions H.-R.L., J.-I.K. and B.-K.K. performed and analysed all the electrophysiological experiments and data; H.Y.G., E.S.J. and J.-S.L. generated and characterized *Shank2*$^{-/-}$ mice; S.-G.P. performed USV experiments; H.W., W.M. and J.L. performed immunoblot analysis; H.W., W.M., S.H. and C.C. contributed to mouse breeding and behavioural characterization; Y.S.C. performed electron microscopy experiments; H.W. and W.M. conducted all the other experiments; K.L., D.K., Y.C.B., B.-K.K., M.G.L. and E.K. supervised the project and wrote the manuscript. B.-K.K., M.G.L. and E.K. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.-K.K. (kaang@snu.ac.kr), M.G.L. (mlee@yuhs.ac) or E.K. (kime@kaist.ac.kr).

LETTER

doi:10.1038/nature11114

The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma

Pedro A. Pérez-Mancera[1], Alistair G. Rust[2], Louise van der Weyden[2], Glen Kristiansen[3], Allen Li[4], Aaron L. Sarver[5], Kevin A. T. Silverstein[5], Robert Grützmann[6], Daniela Aust[7], Petra Rümmele[8], Thomas Knösel[9,10], Colin Herd[11], Derek L. Stemple[11], Ross Kettleborough[11], Jacqueline A. Brosnan[4], Ang Li[4], Richard Morgan[4], Spencer Knight[4], Jun Yu[4], Shane Stegeman[12], Lara S. Collier[13], Jelle J. ten Hoeve[14,15], Jeroen de Ridder[14], Alison P. Klein[4], Michael Goggins[4], Ralph H. Hruban[4], David K. Chang[16,17,18], Andrew V. Biankin[16,17,18], Sean M. Grimmond[19], Australian Pancreatic Cancer Genome Initiative†, Lodewyk F. A. Wessels[14,15], Stephen A. Wood[12], Christine A. Iacobuzio-Donahue[4]*, Christian Pilarsky[6]*, David A. Largaespada[20]*, David J. Adams[2] & David A. Tuveson[1]

Pancreatic ductal adenocarcinoma (PDA) remains a lethal malignancy despite much progress concerning its molecular characterization. PDA tumours harbour four signature somatic mutations[1–4] in addition to numerous lower frequency genetic events of uncertain significance[5]. Here we use *Sleeping Beauty* (SB) transposon-mediated insertional mutagenesis[6,7] in a mouse model of pancreatic ductal preneoplasia[8] to identify genes that cooperate with oncogenic *Kras^{G12D}* to accelerate tumorigenesis and promote progression. Our screen revealed new candidate genes for PDA and confirmed the importance of many genes and pathways previously implicated in human PDA. The most commonly mutated gene was the X-linked deubiquitinase *Usp9x*, which was inactivated in over 50% of the tumours. Although previous work had attributed a pro-survival role to *USP9X* in human neoplasia[9], we found instead that loss of *Usp9x* enhances transformation and protects pancreatic cancer cells from anoikis. Clinically, low USP9X protein and messenger RNA expression in PDA correlates with poor survival after surgery, and USP9X levels are inversely associated with metastatic burden in advanced disease. Furthermore, chromatin modulation with trichostatin A or 5-aza-2′-deoxycytidine elevates *USP9X* expression in human PDA cell lines, indicating a clinical approach for certain patients. The conditional deletion of *Usp9x* cooperated with *Kras^{G12D}* to accelerate pancreatic tumorigenesis in mice, validating their genetic interaction. We propose that *USP9X* is a major tumour suppressor gene with prognostic and therapeutic relevance in PDA.

The biological sequelae of PDA has been partially attributed to frequent and well characterized mutations in *KRAS* (>90%), *CDKN2A* (>90%), *TP53* (70%) and *SMAD4* (55%)[1–4]. Recent genome-wide analyses have uncovered numerous additional somatic genetic alterations, although the functional relevance of most remains uncertain[5]. To explore the molecular genesis of PDA we previously generated a mouse model of pancreatic intraepithelial neoplasia (mPanIN) by conditionally expressing an endogenous *Kras^{G12D}* allele in the developing pancreas[8]. Mice with mPanIN spontaneously progress to mouse PDA (mPDA) after a long and variable latency, providing an opportunity to characterize genes that cooperate with *Kras^{G12D}* to promote early mPDA. We hypothesized that such genes could be directly identified by applying insertional mutagenesis strategies[6,7,10,11] in our mPanIN model, and that these candidates could represent 'drivers' of PDA development.

Accordingly, we interbred our mPanIN model with two distinct SB transposon systems and monitored mice for early disease progression. Our initial approach used the well characterized *CAGGS-SB10* transgenic allele to promote transposition[6]. Although *CAGGS-SB10* promoted PDA, a variety of non-pancreatic neoplasms and a paucity of identified common insertion sites (CIS) in the recovered pancreatic neoplasms precluded a comprehensive analysis, potentially reflecting the variegated expression of *CAGGS-SB10* (ref. 12) (Supplementary Figs 1a and 2, and Supplementary Tables 1 and 3b).

To increase the specificity and potency of SB mutagenesis, we generated a conditional *SB13* mutant mouse by targeting the *Rosa26* locus in embryonic stem cells (Supplementary Fig. 3a, b). The pancreas-specific expression and function of the conditional *SB13* allele was confirmed (Supplementary Fig. 3c), and we found that SB13-induced transposition by itself did not promote lethality or pancreatic tumorigenesis (Fig. 1a and Supplementary Fig. 4a). In contrast, *Kras$^{LSL-G12D}$; Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice rapidly progressed and succumbed to invasive pancreatic neoplasms (Fig. 1a–c). A cohort of 117 *Kras$^{LSL-G12D}$; Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice (Supplementary Fig. 1b) was monitored for tumour development, and 103 of these mice were available for full necropsy and tissue procurement. The majority of such mice harboured multi-focal pancreatic tumours, and 198 distinct primary tumours and metastases were subjected to histological and molecular analysis. Most mice had invasive carcinomas (66 of 103) that consisted of classical mPDA (78.8%) or invasive cystic neoplasms (21.2%); 34.8% of mice also contained metastases predominantly in their liver and lungs (Supplementary Fig. 4c). The remainder of the mice (37 of 103) had pre-invasive pancreatic tumours consisting of high-grade

[1]Li Ka Shing Centre, Cambridge Research Institute, Cancer Research UK, and Department of Oncology, Robinson Way, Cambridge CB2 0RE, UK. [2]Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [3]Institute of Pathology, University Hospital of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany. [4]Departments of Oncology and Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. [5]Biostatistics and Bioinformatics Core, Masonic Cancer Center, University of Minnesota, 425 Delaware St SE MMC 806, Minneapolis, Minnesota 55455, USA. [6]Department of Surgery, University Hospital Dresden, Fetscherstr. 74, 01307 Dresden, Germany. [7]Institute of Pathology, University Hospital Dresden, Fetscherstr. 74, 01307 Dresden, Germany. [8]Institute of Pathology, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany. [9]Institute of Pathology, University Hospital of Jena, Bachstraße 18, 07743 Jena, Germany. [10]Institute of Pathology, Ludwig-Maximilians-University (LMU), Thalkirchnerstr. 36, 80337 Munich, Germany. [11]Vertebrate Development and Genetics, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [12]Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, Queensland 4111, Australia. [13]School of Pharmacy, University of Wisconsin, Madison, Wisconsin 53705, USA. [14]Delft Bioinformatics Lab, Faculty of EEMCS, Delft University of Technology, 2628 CD Delft, The Netherlands. [15]Bioinformatics and Statistics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [16]The Kinghorn Cancer Centre, Cancer Research Program, Garvan Institute of Medical Research, 372 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia. [17]Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia. [18]South Western Sydney Clinical School, Faculty of Medicine, University of NSW, Liverpool, New South Wales 2170, Australia. [19]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. [20]Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA.
*These authors contributed equally to this work.
†Lists of participants and their affiliations appear at the end of the paper.

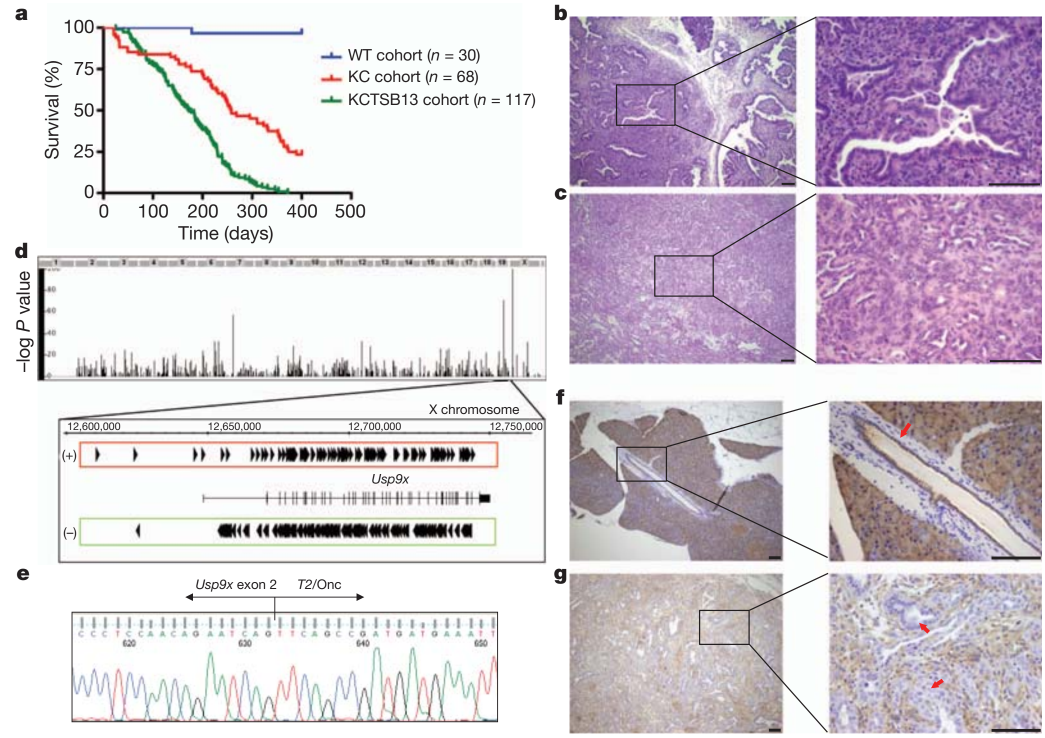


Figure 1 | Transposon mutagenesis accelerates murine PDA and targets Usp9x. **a**, Increased mortality of *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* (KCTSB13) mice compared to the KC cohort (containing *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; T2/Onc*, *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre*; *Rosa26-LSL-SB13* and *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre* mice) (172 versus 257 days, $P < 0.001$; long-rank test). The wild-type (WT) cohort comprises *Kras*$^{LSL\text{-}G12D}$; *T2/Onc; Rosa26-LSL-SB13* and *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice. **b**, **c**, Invasive cystic neoplasm (**b**) and mPDA (**c**) in *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice. Scale bar: 100 μm. **d**, *Usp9x* is the major CIS in *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* PDA tumours (*x* axis denotes genome, *y* axis –log *P* value), with bidirectional insertions. (+) indicates parallel to *Usp9x* expression; (−) indicates antiparallel. **e**, *Usp9x* exon 2-*T2/Onc* chimaeric mRNA in SB13 tumours. **f**, **g**, Usp9x protein expression in normal pancreatic ducts (arrow) (**f**), but not in neoplastic cells (arrows) (**g**), in SB13 PDA harbouring Usp9x insertions. Scale bar: 100 μm.

mPanIN and cyst-forming papillary neoplasms (Supplementary Fig. 4b).

The candidate genes identified from the SB13 screen represented unanticipated candidates as well as many genes and pathways previously implicated in human PDA (Table 1 and Supplementary Tables 2, 3a and 4). Indeed, various members of the TGF-β pathway, including *Smad3*, *Smad4*, *Tgfbr1* and *Tgfbr2*, were collectively mutated in 32% of the tumours. Also, the Rb–p16Ink4a pathway was disrupted in 21% of the tumours. CISs representing the orthologues of additional human PDA genes included *Fbxw7* (24.2%), *Arid1a* (19.1%), *Acvr1b* (19.1%), *Stk11* (also called *Lkb1*) (6.5%), *Mll3* (6%), *Smarca4* (6%) and *Pbrm1* (4.5%)[5,13–15]. *Trp53* was the only commonly mutated PDA gene conspicuously absent, although the p53 regulatory deubiquitinase Usp7 was a CIS (6.5%)[16]. Several CISs previously noted in insertional mutagenesis screens for hepatocellular carcinoma or gastrointestinal tract adenomas, but not typically mutated in PDA, were also identified in this study, including *Zbtb20*, *Nfib* and *Ube2h* in liver tumours[10], and *Pten*, *Tcf12*, *Ppp1r12a* and *Ankrd11* in gastrointestinal tract adenoma/adenocarcinoma[11]. This indicates that many tumour progression pathways may be common to pancreatic, liver and gastrointestinal/colorectal tumours.

Table 1 | Top 20 candidate CIS genes that cooperate with *Kras*^{G12D} to promote mPDA in *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice

Gene	Chr	CIS peak location	CIS height	*n*	*I*	Mutation in humans
Usp9x	X	12691773	158.1266	101	341	–
Pten	19	32872602	64.5204	61	96	–
Fndc3b	3	27562591	13.7096	55	67	–
Setd5	6	113057997	35.6176	52	71	–
Arfip1/Fbxw7	3	84769635	21.6666	48	80	Yes (ref. 5)
Fam193a	5	34705809	24.3555	45	78	–
Ctnna1	18	35342868	20.2017	45	50	–
Magi1	6	93859940	13.3715	43	57	–
Mkln1	6	31414109	16.5263	41	53	–
Pum1	4	130288478	12.7948	41	46	Yes (ref. 5)
Farp1	14	121587858	9.407	39	47	–
Foxp1	6	98921646	19.5831	38	60	–
Arid1a	4	133268936	32.1628	38	47	Yes (ref. 5)
Acvr1b	15	101024934	31.1752	38	47	Yes (ref. 15)
Map4k3	17	81109860	13.2385	38	45	Yes (ref. 5)
Stag2	X	39535994	16.8613	37	48	–
Mll5	5	22982314	16.0001	37	43	Yes (ref. 5)
Atxn2/Sh2b3	5	122267680	12.3174	37	41	–
Arhgap5	12	53644560	37.416	35	61	–
Gsk3b	16	38106972	21.79	35	43	–

CISs were ranked by tumour frequency where the spatial distribution of insertion sites was analysed using the narrowest 15K kernel scale. Chr, chromosome; CIS height, estimate of the number of insertions within a specific genomic region as a result of summing the kernel functions present in that region; *I*, total number of insertions of the CIS in the indicated tumours; *n*, number of tumours from which the CIS was found.

Unexpectedly, the most frequent CIS observed was the X-linked deubiquitinase *Usp9x*, a gene that had not been previously associated with PDA or other types of carcinoma in humans or mouse models. Indeed, the COSMIC data base revealed only one *USP9X* mutation in a case of ovarian cancer, although the functional relevance of this mutation has not been characterized (COSMIC mutation ID 73237). *Usp9x* was disrupted in over 50% of all tumours, with 341 insertions noted in the 101 tumours harbouring this CIS (Fig. 1d and Table 1). Furthermore, *Usp9x* was also identified as a CIS in four samples from the initial SB10 screen (Supplementary Table 1), supporting its candidacy as a PDA genetic determinant. We confirmed that *Usp9x* was disrupted in tumours by isolating chimaeric fusion mRNAs that spliced the *Usp9x* transcript to the *T2/Onc* transposon (Fig. 1e). In addition, the Usp9x protein was specifically absent in neoplastic cells in pancreatic tumours bearing intragenic insertions (Fig. 1f, g).

To characterize the cellular and molecular pathways affected by *Usp9x* in PDA, we used RNA interference to deplete *Usp9x* levels in

mPDA cell lines (Supplementary Fig. 5a). Although *Usp9x* depletion did not affect the proliferation of monolayer cultures (Supplementary Fig. 5b), it significantly increased colony formation in soft agar (Fig. 2a, Supplementary Fig. 5c) compared to cells transfected with scrambled short hairpin RNAs. Furthermore, knock down of *Usp9x* potently suppressed anoikis in mPDA cells (Fig. 2b). These properties of Usp9x were predominantly dependent on its intrinsic deubiquitinase activity (Supplementary Fig. 6a, b).

Because USP9X was previously reported to positively regulate SMAD4 transcriptional activity[17] and *SMAD4* is commonly mutated in PDA[4], we hypothesized that *Usp9x* loss would attenuate Smad4 function or TGF-β responsiveness in PDA cell lines. However, irrespective of Usp9x expression level, mPDA cell lines expressed Smad4 and were equally sensitive to p21 induction, growth inhibition and morphological alterations after exposure to TGF-β1 (Supplementary Fig. 7). Therefore, we were unable to ascribe a specific role to Usp9x in the regulation of the Smad4–TGF-β pathway in mPDA cells or tumours.

We next investigated several additional proteins reported to be regulated by Usp9x and involved in pathways relevant to cellular transformation. Although USP9X has been shown to bind to and regulate two proteins involved in cell survival, ASK1 (ref. 18) and MCL1 (refs 9, 19), we could not detect obvious changes in Ask1 or Mcl1 protein levels upon *Usp9x* loss (Fig. 2c). Usp9x has also been reported to deubiquitinate and thereby stabilize the E3 ligase Itch[20]; decreased protein levels of Itch were observed in mouse and human PDA cells upon the depletion of Usp9x (Fig. 2c and Supplementary Fig. 8a). Notably, ectopic Itch expression was sufficient to promote anoikis in mPDA cells (Fig. 2d), and Itch was partially responsible for the ability of Usp9x to promote anoikis and suppress colony formation (Supplementary Fig. 6c, d). Because Itch is known to promote the degradation of several proteins relevant to cell proliferation and survival[21], we evaluated the protein expression of likely candidates

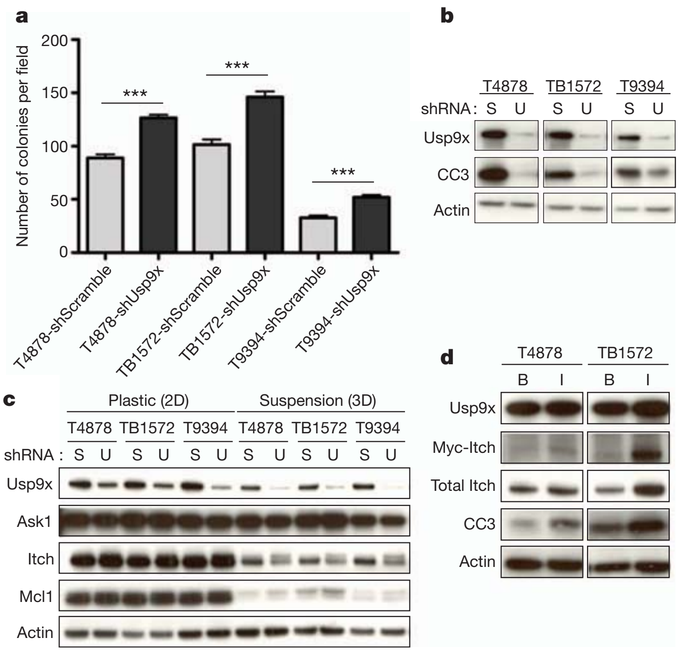


Figure 2 | Usp9x regulates PDA cellular transformation and Itch. **a**, **b**, Usp9x knock down promotes anchorage-independent growth in three mPDA cell lines (**a**), and decreases anoikis denoted by cleaved caspase 3 (CC3) (**b**). The mean and s.e.m. of one representative experiment performed in triplicate are shown in **a** (***$P < 0.001$; Mann–Whitney *U*-test). S, scramble; U, Usp9x. **c**, Usp9x knock down decreases Itch levels but not Ask1 or Mcl1 levels. Changes in Itch are more evident in suspension cultures, and the slower migrating band has the expected mobility of mono-ubiquitinated Itch. **d**, Ectopic Itch induces anoikis. B, pBabe-neo; I, pBabe-neo-Myc-Itch.

including c-Jun, p63 and c-Flip but observed no alterations (Supplementary Fig. 8b). Furthermore, the *Itch* gene was identified as a CIS in 13% of cases (Supplementary Table 2). Therefore, Usp9x mutation may promote tumorigenesis in part by disabling Itch function, and the Usp9x–Itch pathway may work to constrain pancreatic tumorigenesis.

To determine whether *USP9X* expression is aberrant in human PDA, three distinct patient cohorts were assessed. First, we analysed a cohort of 100 Australian patients who underwent surgery for localized PDA and had detailed information available concerning clinical-pathological characteristics and outcome (Supplementary Fig. 9 and Supplementary Tables 5 and 6). Tumour DNA from 88 patients in this cohort failed to yield somatic mutations in *USP9X*, consistent with previous reports[5] (data not shown). Notably, the low expression of *USP9X* mRNA correlated with poor survival after surgery ($P = 0.0076$) (Fig. 3a), and multivariate analysis revealed that *USP9X* expression was an independent poor prognostic factor after surgery (Supplementary Table 7). We next analysed autopsy specimens from a separate cohort of 42 American patients to determine that USP9X protein expression inversely correlated with a widespread metastatic pattern ($P = 0.0212$) (Fig. 3b), and bore no relation to SMAD4 expression (Supplementary Table 8). A third collection of PDA specimens obtained from resected German patients ($n = 404$) was used to determine that USP9X and ITCH protein levels were decreased (Supplementary Fig. 10) and expressed in a similar manner (Spearman-Rho correlation = 0.47; $P < 0.01$; Supplementary Table 9a) in tumours compared to normal pancreatic tissue. Furthermore, the proportion of tumours that had undetectable USP9X (13.6%) or ITCH (30.5%) protein correlated with a worse outcome (Supplementary Fig. 11, Supplementary Table 9b, c), particularly regarding USP9X in the subset of high-grade tumours (Fig. 3c and Supplementary Tables 10 and 11). Collectively, these findings implicate the loss of USP9X expression as a relevant event in human pancreatic cancer progression.

We found that USP9X was expressed throughout murine and human tumour development and lost focally in PDAs (Supplementary Figs 12 and 13). Additionally, human PDA cell lines expressed lower levels of USP9X compared to non-PDA cancer cell lines (Supplementary Fig. 14). To investigate additional potential mechanisms of *USP9X* regulation in PDA, human cell lines were treated with the DNA methylase inhibitor 5-aza-2′-deoxycytidine and the HDAC inhibitor trichostatin A. Both inhibitors modestly increased the USP9X mRNA and protein levels in most cell lines, indicating that *USP9X* may be epigenetically silenced *in vivo* (Fig. 3d and Supplementary Fig. 15). Furthermore, although the promoter region of *USP9X* was not heavily methylated in tumour samples or PDA cells harbouring low protein expression (data not shown), treatment with 5-aza-2′-deoxycytidine did decrease colony formation of human PDA cells and this was partially reversed by concomitantly knocking down USP9X (Supplementary Fig. 16).

To confirm that *Kras*^{G12D} cooperated with *Usp9x* loss to promote pancreatic cancer, a conditional *Usp9x*^{fl} allele was generated (Supplementary Fig. 17a) and interbred with *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre* mice to evaluate the impact on mPanIN progression. The mosaic expression of Usp9x in pancreas from *Pdx1-cre; Usp9x*$^{fl/y}$ mice was confirmed by immunohistochemistry (Supplementary Fig. 17b). We found that all hemizygous male mice and heterozygous female mice carriers of the *Usp9x*^{fl} allele in the background of *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre* rapidly developed advanced mPanIN and microinvasive neoplasms within 3 months of age (Fig. 3e, f and Supplementary Fig. 18). Immunohistochemical analysis of mPanINs from heterozygous female mice demonstrated absence of Usp9x expression in the pre-neoplastic and neoplastic cells (Supplementary Fig. 18), indicative of additional events such as X inactivation of the other locus in female mice[22,23]. mPanINs in *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; Usp9x*^{fl} mice expressed intranuclear Smad4, similar to *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre* mice (Supplementary Fig. 19a). Additionally, early passage pancreatic cell cultures

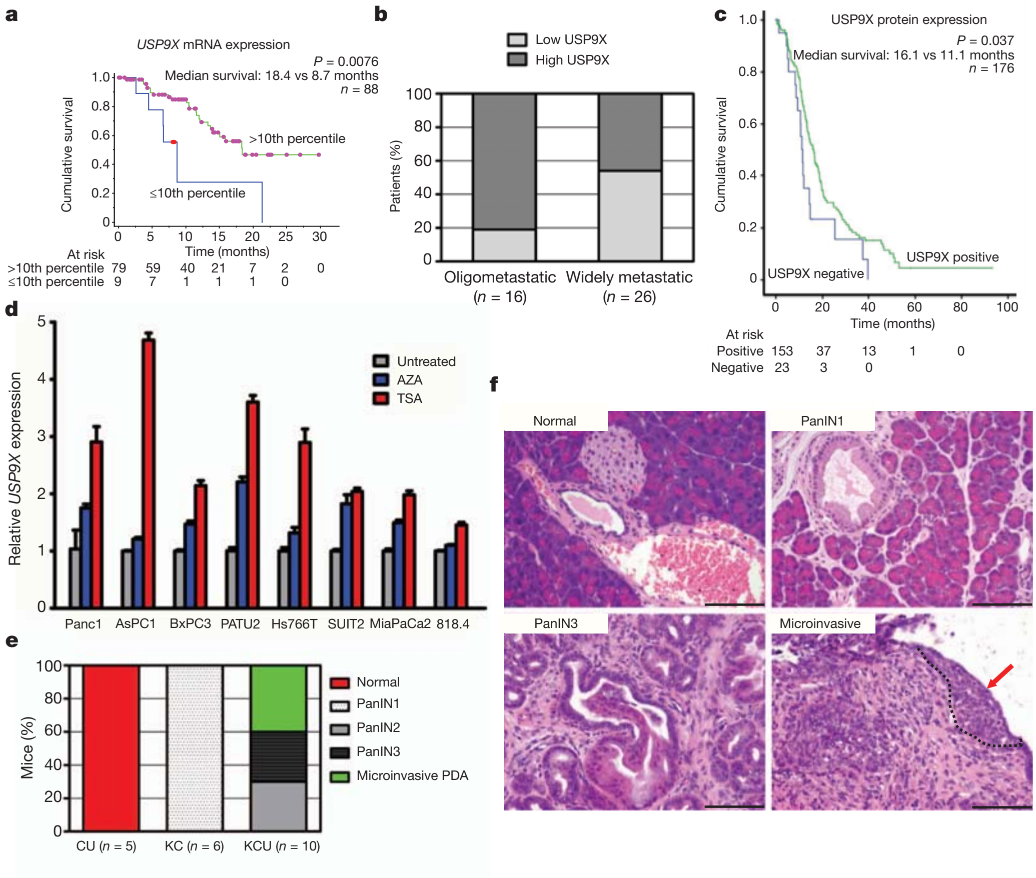


Figure 3 | *USP9X* loss promotes PDA. a–c, Decreased USP9X expression correlates with shortened survival in an Australian post-surgical cohort (**a**) (8.7 versus 18.4 months, $P = 0.0076$; log-rank test), increased metastatic burden in an American autopsy series (**b**) (54% versus 19%, $P = 0.0212$; Fisher's exact test), and diminished survival in a German post-surgical cohort (**c**) (11.1 versus 16.1 months, $P = 0.037$; log-rank test). **d**, Trichostatin A (TSA, red) and 5-aza-2′-deoxycytidine (AZA, blue) modestly increase *USP9X* mRNA expression in a panel of eight human PDA cell lines. The mean and s.e.m. of one representative experiment performed in triplicate are shown. **e**, *Usp9x* deletion promotes mPanIN progression in *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; Usp9x*$^{fl/+}$ and *Kras*$^{LSL\text{-}G12D}$; *Pdx1-cre; Usp9x*$^{fl/y}$ (KCU) mice ($P < 0.0001$; Fisher's exact test). **f**, Representative normal pancreas (*Pdx1-cre; Usp9x*$^{fl/y}$; CU), mPanIN1 (*Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre*; KC), mPanIN3 (*Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; Usp9x*$^{fl/+}$ and *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; Usp9x*$^{fl/y}$; KCU) and microinvasive mPDA (KCU, arrow, circled). Scale bar: 100 µm.

prepared from these mice confirmed the absence of the Usp9x protein and altered regulation of Itch (Supplementary Fig. 19b). Although some mice died of local or metastatic pancreatic cancer, aggressive oral papillomas often required the culling of young mice and demonstrated that *Kras*^{G12D} and *Usp9x* loss also cooperated to transform keratinocytes (Supplementary Fig. 19c).

Although a recent report implicated *USP9X* as a pro-survival gene by stabilizing MCL1 (ref. 9), potential inhibitors of USP9X should be developed with caution as we find that *Usp9x* has tissue-specific effects including a tumour suppressor role in oncogenic *Kras*-initiated pancreatic carcinoma. *USP9X* is probably epigenetically silenced in a subset of PDA, thus explaining why previous DNA sequencing efforts have failed to identify this as a participant in carcinogenesis, and indicating that clinically available epigenome modulators may be beneficial agents in such patients. ITCH is a likely mediator of pancreatic tumour suppression by USP9X, and continued investigation of the USP9X–ITCH pathway is warranted. More generally, the identification of *Usp9x* through the use of transposon mutagenesis reaffirms the utility of *in vivo* mouse cancer screens to complement the direct investigation of human cancer.

METHODS SUMMARY

Kras$^{LSL\text{-}G12D}$ (ref. 24), *Pdx1-cre* (ref. 8), *T2/Onc* (ref. 6), *CAGGS-SB10* (ref. 6) and *Rosa26-LSL-SB13* strains were interbred to generate *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre*, *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; T2/Onc; CAGGS-SB10* and *Kras*$^{LSL\text{-}G12D}$*; Pdx-1-cre*; *T2/Onc*; *Rosa26-LSL-SB13* compound mutant mice. Non-quadruple mutant mice represented the comparison cohorts. *Kras*$^{LSL\text{-}G12D}$ and *Pdx1-cre* mice were interbred with *Usp9x*^{fl} mice to generate the *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; Usp9x*$^{fl/+}$ and *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre; Usp9x*$^{fl/y}$ compound mutant mice, as well as the two control cohorts *Pdx1-cre; Usp9x*$^{fl/y}$ and *Kras*$^{LSL\text{-}G12D}$*; Pdx1-cre*. *Usp9x*^{fl} mice were generated by Ozgene Pty. Ltd. Mice were maintained in compliance with the UK home office regulations. Splinkerette PCRs were performed as described previously[25,26]. Reads from sequenced tumours were mapped to the mouse genome assembly NCBI m37 and merged together to identify SB insertion sites, as previously described[25]. Redundant sequences, as well as insertions in the *En2* gene and in the *T2/Onc* donor concatemer resident chromosome (chromosome 1), were removed. Mouse survival curves and cell culture experiments were analysed with the GraphPad prism program. The IHC histoscoring from the TMA samples and Kaplan–Meier survival curves were analysed with SPSS18, and the Spearman-Rho correlation coefficient (two-sided) between USP9X and ITCH was calculated. The IHC USP9X histoscore and analysis was conducted using Fisher's exact test on post-mortem samples.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May 2011; accepted 5 April 2012.
Published online 29 April 2012.

1. Almoguera, C. *et al.* Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell* **53,** 549–554 (1988).
2. Caldas, C. *et al.* Frequent somatic mutations and homozygous deletions of the p16 (*MTS1*) gene in pancreatic adenocarcinoma. *Nature Genet.* **8,** 27–32 (1994).
3. Redston, M. S. *et al.* p53 mutations in pancreatic carcinoma and evidence of common involvement of homocopolymer tracts in DNA microdeletions. *Cancer Res.* **54,** 3025–3033 (1994).

4. Hahn, S. A. *et al.* DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271,** 350–353 (1996).
5. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–1806 (2008).
6. Collier, L. S., Carlson, C. M., Ravimohan, S., Dupuy, A. J. & Largaespada, D. A. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436,** 272–276 (2005).
7. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* **436,** 221–226 (2005).
8. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4,** 437–450 (2003).
9. Schwickart, M. *et al.* Deubiquitinase USP9X stabilizes MCL1 and promotes tumour cell survival. *Nature* **463,** 103–107 (2010).
10. Keng, V. W. *et al.* A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nature Biotechnol.* **27,** 264–274 (2009).
11. Starr, T. K. *et al.* A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. *Science* **323,** 1747–1750 (2009).
12. Collier, L. S. *et al.* Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. *Cancer Res.* **69,** 8429–8437 (2009).
13. Avizienyte, E. *et al.* LKB1 somatic mutations in sporadic tumors. *Am. J. Pathol.* **154,** 677–681 (1999).
14. Varela, I. *et al.* Exome sequencing identifies frequent mutation of the SWI/SNF complex gene *PBRM1* in renal carcinoma. *Nature* **469,** 539–542 (2011).
15. Su, G. H. *et al.* *ACVR1B* (*ALK4*, activin receptor type 1B) gene mutations in pancreatic carcinoma. *Proc. Natl Acad. Sci. USA* **98,** 3254–3257 (2001).
16. Li, M. *et al.* Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* **416,** 648–653 (2002).
17. Dupont, S. *et al.* FAM/USP9x, a deubiquitinating enzyme essential for TGFβ signaling, controls Smad4 monoubiquitination. *Cell* **136,** 123–135 (2009).
18. Nagai, H. *et al.* Ubiquitin-like sequence in ASK1 plays critical roles in the recognition and stabilization by USP9X and oxidative stress-induced cell death. *Mol. Cell* **36,** 805–818 (2009).
19. Sun, H. *et al.* Bcr-Abl ubiquitination and Usp9x inhibition block kinase signaling and promote CML cell apoptosis. *Blood* **117,** 3151–3162 (2011).
20. Mouchantaf, R. *et al.* The ubiquitin ligase itch is auto-ubiquitylated *in vivo* and *in vitro* but is protected from degradation by interacting with the deubiquitylating enzyme FAM/USP9X. *J. Biol. Chem.* **281,** 38738–38747 (2006).
21. Bernassola, F., Karin, M., Ciechanover, A. & Melino, G. The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell* **14,** 10–21 (2008).
22. Wang, X., Soloway, P. D. & Clark, A. G. Paternally biased X inactivation in mouse neonatal brain. *Genome Biol.* **11,** R79 (2010).
23. Yang, F., Babak, T., Shendure, J. & Disteche, C. M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* **20,** 614–622 (2010).
24. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15,** 3243–3248 (2001).
25. March, H. N. *et al.* Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nature Genet.* **43,** 1202–1209 (2011).
26. Uren, A. G. *et al.* A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nature Protocols* **4,** 789–798 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Labosky for assistance in generating the *Rosa26-LSL-SB13* mouse; B. Bhagavan for pathology consultation; M. Tsao for providing the HPDE cell line; and N. Copeland and K. Mann for sharing pre-published information. We thank A. Gopinathan, H. Tiriac, D. Engle, D. Chan, F. Connor, S. Derkits and other members of the Tuveson laboratory for assistance and advice, and the animal care staff and histology core at CRI, and The University of Minnesota's Mouse Genetics Laboratory. This research was supported by the University of Cambridge and Cancer Research UK, The Li Ka Shing Foundation and Hutchison Whampoa Limited, the NIHR Cambridge Biomedical Research Centre, and the NIH (2P50CA101955 SPORE grant to D.A.T., D.A.L. and C.A.I.-D.; grants CA62924, CA128920 and CA106610 to C.A.I.-D.; P50CA62924 SPORE grant to R.H.H. and C.A.I.-D.; and CA122183 to L.S.C.). D.J.A. is supported by Cancer Research UK and the Wellcome Trust. L.v.d.W. is supported by the Kay Kendall Leukemia Fund. C.P. is supported by Wilhelm Sander Stiftung (2009.039.1) and Deutsche Forschungsgemeinschaft (PI 341/5-1). A.V.B., D.K.C., S.M.G. and the APGI investigators are funded by the National Health and Medical research Council of Australia (NHMRC); Queensland Government; Cancer Council NSW; Australian Cancer Research Foundation; Cancer Institute NSW; The Avner Nahmani Pancreatic Cancer Research Foundation; and the R.T. Hall Trust. Additional support was obtained from Fundación Ibercaja (P.A.P.-M.).

Author Contributions P.A.P.-M. performed the majority of all experiments, designed experiments, analysed data, and wrote the manuscript. L.v.d.W. and J.A.B. performed *in vitro* experiments. S.S. and S.A.W. generated the conditional Usp9x mouse. L.S.C. provided the *CAGGS-SB10* and *T2/Onc* mice. A.G.R., A.L.S., K.A.T.S., J.J.t.H., J.d.R. and L.F.A.W. conducted the CIS data analysis. G.K., R.G., D.A., P.R., T.K. and C.P. generated data from resected pancreatic tumours. Allen Li, R.H.H., R.M., S.K., J.Y., Ang Li, M.G. and C.A.I.-D. analysed human samples from autopsy series, and analysed mouse pathology and methylation studies. C.H., D.L.S. and R.K. sequenced PDA human samples from autopsy series. A.P.K. provided statistical analyses for the human PDA data sets. APGI, D.K.C., S.M.G. and A.V.B. generated and analysed data from ICGC/APGI (International Cancer Genome Consortium/Australian Pancreatic Cancer Genome Initiative). D.A.L. provided the *CAGGS-SB10* and *T2/Onc* mice, and analysed data. D.J.A. and D.A.T. designed the study, analysed the data, and wrote the manuscript. All authors commented upon and edited the final manuscript.

Author Information The GEO accession number for the ICGC/APGI gene expression data is GSE36924. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.J.A. (da1@sanger.ac.uk) or D.A.T. (david.tuveson@cancer.org.uk).

Australian Pancreatic Cancer Genome Initiative

Garvan Institute of Medical Research Andrew V. Biankin[1], Amber L. Johns[1], Amanda Mawson[1], David K. Chang[1], Mary-Anne L. Brancato[1], Sarah J. Rowe[1], Skye L. Simpson[1], Mona Martyn-Smith[1], Lorraine A. Chantrill[1], Venessa T. Chin[1], Angela Chou[1], Mark J. Cowley[1], Jeremy L. Humphris[1], Marc D. Jones[1], R. Scott Mead[1], Adnan M. Nagrial[1], Marina Pajic[1], Jessica Pettit[1], Mark Pinese[1], Ilse Rooman[1], Jianmin Wu[1], Roger J. Daly[1], Elizabeth A. Musgrove[1], Robert L. Sutherland[1]; **Queensland Center for Medical Genomics, Institute for Molecular Bioscience** Sean M. Grimmond[2], Nicola Waddell[2], Karin S. Kassahn[2], David K. Miller[2], Peter J. Wilson[2], Ann-Marie Patch[2], Sarah Song[2], Ivon Harliwong[2], Senel Idrisoglu[2], Craig Nourse[2], Ehsan Nourbakhsh[2], Suzanne Manning[2], Shivangi Wani[2], Milena Gongora[2], Matthew Anderson[2], Oliver Holmes[2], Conrad Leonard[2], Darrin Taylor[2], Scott Wood[2], Christina Xu[2], Katia Nones[2], J. Lynn Fink[2], Angelika Christ[2], Tim Bruxner[2], Nicole Cloonan[2], Felicity Newell[2], John V. Pearson[2]; **Royal North Shore Hospital** Jaswinder S. Samra[3], Anthony J. Gill[3], Nick Pavlakis[3], Alex Guminski[3], Christopher Toon[3]; **Bankstown Hospital** Andrew V. Blankin[4], Ray Asghari[4], Neil D. Merrett[4], David K. Chang[4], Darren A. Pavey[4], Amithabad Das[4]; **Liverpool Hospital** Peter H. Cosman[5], Kasim Ismail[5], Chelsie O'Connor[5]; **Westmead Hospital** Vincent W. Lam[6], Duncan McLeod[6], Henry C. Pleass[6], Virginia James[6]; **Royal Prince Alfred Hospital** James G. Kench[7], Caroline L. Cooper[7], David Joseph[7], Charbel Sandroussi[7], Michael Crawford[7]; **Fremantle Hospital** Michael Texler[8], Cindy Forrest[8], Andrew Laycock[8], Krishna P. Epari[8], Mo Ballal[8], David R. Fletcher[8], Sanjay Mukhedkar[8]; **Sir Charles Gairdner Hospital** Nigel A. Spry[9], Bastiaan DeBoer[9], Ming Chai[9], Kynan Feeney[9]; **St John of God Healthcare** Nikolajs Zeps[10], Maria Beilin[10]; **Royal Adelaide Hospital** Nam Q. Nguyen[11], Andrew R. Ruszkiewicz[11], Chris Worthley[11], Chuan P. Tan[11], Tamara Debrencini[11]; **Flinders Medical Center** John Chen[12], Mark E. Brooke-Smith[12], Virginia Papangelis[12]; **Greenslopes Private Hospital** Henry Tang[13], Andrew P. Barbour[13]; **Envoi Pathology** Andrew D. Clouston[14], Patrick Martin[14]; **Princess Alexanda Hospital** Thomas J. O'Rourke[15], Amy Chiang[15], Jonathan W. Fawcett[15], Kellee Slater[15], Shinn Yeung[15], Michael Hatzifotis[15], Peter Hodgkinson[15]; **Austin Hospital** Christopher Christophi[16], Mehrdad Nikfarjam[16], Victorian Cancer Biobank[17]; **Johns Hopkins Medical Institutes** James R. Eshleman[18], Ralph H. Hruban[18], Anirban Maitra[18], Christine A. Iacobuzio-Donahue[18], Richard D. Schulick[18], Christopher L. Wolfgang[18], Richard A. Morgan[18]; **ARC-NET Center for Applied Research on Cancer** Rita T. Lawlor[19], Stefania Beghelli[19], Vincenzo Corbo[19], Maria Scardoni[19], Claudio Bassi[19]; **University of California** Margaret A. Tempero[20]

Australian Pancreatic Cancer Genome Initiative

[1]The Kinghorn Cancer Centre, Garvan Institute of Medical Research, 372 Victoria Street, Darlinghurst, Sydney, New South Wales 2010, Australia. [2]Queensland Center for Medical Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. [3]Royal North Shore Hospital, Westbourne Street, Saint Leonards, New South Wales 2065, Australia. [4]Bankstown Hospital, Eldridge Road, Bankstown, New South Wales 2200, Australia. [5]Liverpool Hospital, Elizabeth Street, Liverpool, New South Wales 2170, Australia. [6]Westmead Hospital, Corner of Hawkesbury and Darcy Roads, Westmead, New South Wales 2145, Australia. [7]Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050, Australia. [8]Fremantle Hospital, Alma Street, Fremantle, Western Australia 6959, Australia. [9]Sir Charles Gardiner Hospital, Hospital Avenue, Nedlands, Western Australia 6009, Australia. [10]St John of God Healthcare, 12 Salvado Road, Subiaco, Western Australia 6008, Australia. [11]Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. [12]Flinders Medical Center, Flinders Drive, Bedford Park, South Australia 5042, Australia. [13]Greenslopes Private Hospital, Newdegate Street, Greenslopes, Queensland 4120, Australia. [14]Envoi Pathology, 1/49 Butterfield Street, Herston, Queensland 4006, Australia. [15]Princess Alexandria Hospital, Corner of Cornwall Street and Ipswich Road, Woolloongabba, Queensland 4102, Australia. [16]Austin Hospital, 145 Studley Road, Heidelberg, Victoria 3084, Australia. [17]Victorian Cancer Biobank, 1 Rathdowne Street, Carlton, Victoria 3053, Australia. [18]Johns Hopkins Medical Institutes, 600 North Wolfe Street, Baltimore, Maryland 21287, USA. [19]ARC-NET Center for Applied Research on Cancer, University of Verona, Via dell'Artigliere 19, 37129 Verona, Province of Verona, Italy. [20]University of California, San Francisco, 500 Parnassus Avenue, San Francisco, California 94122, USA.

METHODS

Mouse strains. *Kras*$^{LSL-G12D}$ (ref. 24), *Pdx1-cre* (ref. 8), *T2/Onc* (ref. 6), *CAGGS-SB10* (ref. 6) and *Rosa26-LSL-SB13* strains were interbred to generate *Kras*$^{LSL-G12D}$; *Pdx1-cre*; *T2/Onc*; *CAGGS-SB10* (KCTSB10) and *Kras*$^{LSL-G12D}$; *Pdx-1-cre*; *T2/Onc*; *Rosa26-LSL-SB13* (KCTSB13) compound mutant mice. Non-quadruple mutant mice represented the comparison cohorts. Genomic DNA from tumours developed in KCTSB10 and KCTSB13 mice was obtained using the Puregene Core Kit A (Qiagen) and splinkerette PCRs were performed as described previously[25,26]. For the KCU cohort, *Kras*$^{LSL-G12D}$ and *Pdx1-cre* mice were interbred with *Usp9x*^{fl} mice to generate the *Kras*$^{LSL-G12D}$; *Pdx1-cre*; *Usp9x*$^{fl/+}$ and *Kras*$^{LSL-G12D}$; *Pdx1-cre; Usp9x*$^{fl/y}$ (KCU) compound mutant mice, as well as the two control cohorts *Pdx1-cre; Usp9x*$^{fl/y}$ (CU) and *Kras*$^{LSL-G12D}$; *Pdx1-cre* (KC).

Generation of *Rosa26-LSL-SB13* knock-in mice. TL1 ES cells[27] were electroporated with linearized pRosa26-LSL-SA-SB13-BGHpolyA targeting construct and correctly targeted puromycin-resistant clones were identified by Southern blot. Two positives clones exhibiting a normal karyotype were used to generate chimaeric mice by microinjection into C57BL/6 blastocysts. Germline transmission of the targeted allele was confirmed by Southern blot analysis of tail DNA from the agouti offspring.

***T2/Onc* excision PCR.** Genomic DNAs were obtained from *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* and *T2/Onc; Rosa26-LSL-SB13* mice and primers used to assess the excision of the *T2/Onc* concatemer in the *Pdx1-cre; T2/Onc; Rosa26-LSL-SB13* mice were 5′-TGTGCTGCAAGGCGATTA-3′ and 5′-ACCATGATTACGCC AAGC-3′.

CIS analysis. For the statistical analysis, redundant sequences, as well as insertions in the *En2* gene and in the *T2/Onc* donor concatemer resident chromosome (chromosome 1), were removed and 90,007 non-redundant insertion sites (Supplementary Table 3) were used to identify CISs using a Gaussian kernel convolution framework (GKC)[28]. Reads from sequenced tumours were mapped to the mouse genome assembly NCBI m37 and merged together to identify SB insertion sites, as previously described[25]. An enhanced version of the framework was developed for SB screens to account for the local density of TA sites within the genome[25]. For example, a genomic region containing a large number of insertion sites but a low density of TA sites is considered to be significant and thereby identified as a candidate CIS. Conversely, a region with a large number of insertion sites but also containing a high density of TA sites is determined to be less significant, as the transposons have more 'target' sites into which they can integrate. Multiple kernel scales were used in the GKC framework (widths of 15K, 30K, 50K, 75K, 120K and 240K nucleotides). CISs predicted across multiple scales and overlapping in their genomic locations were clustered together, such that the CIS with the smallest genomic 'footprint' was reported as the representative CIS. For highly significant CISs with narrow spatial distributions of insertion sites, the 15K kernel is typically the scale on which CISs are identified. Additional statistical analysis of insertion sites was performed using a Monte Carlo framework[10]. CISs were compared to previously published data sets of human pancreatic cancer genetics[5,29,30].

Detection of *Usp9x-T2/Onc* fusion mRNA by RT–PCR in SB tumours. Total RNA was extracted from snap-frozen SB tumours using the RNeasy Mini kit (Qiagen), and total RNA (1 μg) was reverse transcribed into cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems). RT–PCR was carried out with a nested PCR approach using primers of mouse *Usp9x* exon 1 and the Carp-β-actin splice acceptor sequence of the *T2/Onc* transposon cassette. cDNA was used as a template in a first round of PCR using specific primers corresponding to exon 1 of *Usp9x* (5′-GAGTCTGCGCTGCCGCTGCTG-3′) and Carp-β-actin splice acceptor sequence (5′-CATACCGGCTACGTTGCTAA-3′). The product of this reaction was used as a template in a second round of nested PCR using an internal primer in the *Usp9x* exon 1 (5′-GCTGCCGCTG CTGTTGCTGC-3′) and a second primer in the Carp-β-actin splice acceptor sequence (5′-ACGTTGCTAACAACCAGTGC-3′). PCR products were cloned into pCR 2.1-TOPO vector (Invitrogen) and positives clones sequenced.

Plasmids, shRNAs and transfections. pSuperRetro-PURO retroviral vector (Oligoengine) expressed a short hairpin against mouse and human *USP9X* (5′-GATGAGGAACCTGCATTTC-3′), mouse *Itch* (5′-GACCTGAGAAGACG TTTGT-3′)[31] and a scrambled sequence (5′-GCGCGCTTTGTAGGATTCG-3′). pBabe-zeo-Ecotropic receptor (ecoR) was obtained from Addgene (plasmid no. 10687). Myc-mItch cDNA was released from pCINeo-Myc-Itch (Addgene plasmid no. 11427), and was subcloned in the retroviral vector pBabe-neo (Addgene plasmid no. 1767). KCU1 and KCU2 cell lines were transfected with pEF-DEST51-mUsp9x(WT)-V5 and pEF-DEST51-mUsp9x(C1566S)-V5 plasmids[32,33]. The plasmid pEF/GW-51/LacZ (Invitrogen) was used as control. Transfections were done using Lipofectamine 2000 (Invitrogen). Twenty-four hours later, cells were selected with 5 μg ml^{-1} blasticidin (Invitrogen).

Cell culture. Tumour pancreatic cancer cell lines were established from *Kras*$^{LSL-G12D}$; *Pdx1-cre* (T4878 and T9394), *Kras*$^{LSL-G12D}$; *P48-cre* (TB1572) and *Kras*$^{LSL-G12D}$; *Pdx-1-Cre; Usp9x*^{fl} (KCU1 and KCU2) mice as described previously[34]. Cells were subsequently cultured in DMEM (Invitrogen), supplemented with 10% FCS (Hyclone). The normal human pancreatic ductal cell line HPDE was provided by M. Tsao and cultured as described previously[35,36]. The human pancreatic cancer cell lines AsPC1 (CRL-1682) and BxPC3 (CRL-1687) were acquired from ATCC and cultured according to instructions. The other cell lines were obtained from Clare Hall Laboratories (CRUK). The human cell lines Panc1, MiaPaCa2, 818.4, Hs766T, PATU2, SUIT2, FA6 and MDA-Panc3 (PDA); CaCO2 and SW1116 (colorectal cancer); SKBR3 (breast cancer) and A549 (lung cancer) were cultured in DMEM supplemented with 10% FCS. The human cell lines U937 (histiocytic lymphoma), RAMOS (Burkitt's lymphoma), NCI-H2179 (lung cancer) and ZR75-1 (breast cancer) were cultured in RPMI (Invitrogen) supplemented with 10% FCS. Cells were treated with 1 μM trichostatin A (Sigma) for 24 h or with 5 μM 5-aza-2′-deoxycytidine (Sigma) for 96 h where indicated to obtain RNA and protein lysates to assess USP9X expression. For anchorage-independent growth assay, cells were treated with 5 μM 5-aza-2′-deoxycytidine (Sigma).

Retroviral infections. Phoenix cells were plated 24 h before transfection using the ProFection Mammalian Transfection System Calcium Phosphate (Promega). Target cells were infected with retroviruses produced in the Phoenix packaging cells (24 and 48 h after transfection) in the presence of 8 μg ml^{-1} polybrene (Sigma) and were selected with 2 μg ml^{-1} puromycin (Sigma) or 1 mg ml^{-1} G418 (Clontech). Experiments were performed using at least two independent cell line infected pools. Human PDA cells lines Panc1, SUIT2 and PATU2 infected with retroviral vectors expressed the ecotropic receptor (ecoR).

Transformation, anoikis and EMT assays. Cell lines (1.5×10^4 cells) were plated in triplicate in 12-well plates and counted as indicated using a Z2 Coulter (Beckman). Cells were fed every other day. Anchorage-independent growth was assessed by colony formation in soft agar. Briefly, 15,000 cells were plated in duplicate in DMEM with 15% serum and 0.34% low-melting-point agarose (LMP, BioGene) onto 6-cm dishes coated with 0.5% LMP. Cells were fed twice a week and grown for 2 weeks. Colonies were counted in nine different ×20 fields. For the anoikis assay, 10^5 cells per 0.5 ml were plated in 24-well ultra-low cluster plates (Costar) to allow them to grow in suspension for 4 days. Cells were collected, washed with cold PBS and protein lysates were obtained. Cell line T4878 was cultured in matrigel as previously described[37], plating 1,000 cells per well. Cells were fed every 2 days and grown for 4 days. Epithelial-to-mesenchymal transition (EMT) was determined by plating 10^5 cells per 6-well plates for 24 h to allow attachment, followed by treatment with human TGF-β1 (5 ng ml^{-1}) (RD Systems) for 24 h. p21 induction was assessed after treatment with human TGF-β1 (5 ng ml^{-1}) (RD Systems) for 2 h.

Real-time PCR. Total RNA from human PDA cell lines was extracted using the RNeasy Mini Kit (Qiagen), and total RNA (1 μg) was reverse transcribed into cDNA using the High Capacity RNA-to-cDNA kit (Applied Biosystems). Human *USP9X* expression was analysed by quantitative PCR (qPCR) using TaqMan gene expression assays Hs00245009_m1 (Applied Biosystems) on a 7900HT Real-Time PCR system (Applied Biosystems). Gene expression was normalized to human *ACTB* expression, assessed with the gene expression assays Hs99999903_m1 (Applied Biosystems), and shown relative to control samples.

Western blot analysis. Cells were washed three times in cold PBS and lysed with boiling lysis buffer (1% SDS, 10 mM, pH 7.5 Tris, 50 mM NaF, 1 mM Na_3VO_4). Lysates were boiled for 5 min, passed through a 26 gauge needle to shear genomic DNA and centrifuged for 10 min at 14,000 r.p.m. Equivalent amounts of protein were resolved in 4–12% gradient SDS–PAGE gels (Invitrogen), transferred to Immobilon–P transfer membranes (Millipore), and incubated with the corresponding antibodies including anti-Ask1 (NB110-55482, Novus Biologicals), anti-Mcl1 (5453, Cell Signaling), anti-Usp9x (A301-351A, Bethyl), anti-CC3 (9664, Cell Signaling), anti-Itch (611198, BD), anti-p21 (sc-6246, Santa Cruz), anti-Smad4 (sc-7966, Santa Cruz), anti-Myc tag (2276, Cell Signaling), anti V5 tag (R960-25, Invitrogen), anti-c-Flip (ALX-804-127, Enzo Life Sciences), anti-c-Jun (9165, Cell Signaling), anti-p63 (Ab110038, Abcam), anti-α-Tubulin (T6074, Sigma) and anti-Actin (sc-1616, Santa Cruz Biotechnology). Reactive bands were visualized with ECL plus reagent (Amersham). Relative expression was quantified with Image Quant TL software (GE Healthcare).

Immunohistochemistry. Formalin-fixed paraffin-embedded (FFPE) mouse tissues were cut into 3-μm tissue sections, and antigen retrieval was performed in 10 mM, pH 6.0 citric acid (for Usp9x and E-cadherin) or 10 mM, pH 8.0 EDTA (for Smad4). Endogenous peroxidases were quenched in 3% H_2O_2/PBS for 20 min. Signal detection for immunohistochemistry was accomplished with biotinylated secondary antibodies (Vector Laboratories) using the Elite Vectastain ABC kit and peroxidase substrate DAB kit (Vector Laboratories). Primary antibodies used were anti-Usp9x, 1:200 (A301-351A, Bethyl); E-cadherin, 1:200 (610182, BD) and

anti-Smad4, 1:100 (sc-7966, Santa Cruz). Slides were counterstained with haematoxylin.

Clinical patient samples immunohistochemistry and analysis. Tissue microarrays ($n = 404$) were prepared from patient samples obtained after appropriate informed consent in Dresden (Institute of Pathology, University Hospital Dresden), Regensburg (Institute of Pathology, University Hospital Regensburg) and Jena (Institute of Pathology, University Hospital Jena). Informed consent was obtained for each patient, following review by the human ethics committee Ethikkommission an der Technischen Universität Dresden. The PDA tumour samples were collected from 1993 to 2009, and most of the patients (65%) did not undergo adjuvant chemotherapy. Those that did undergo adjuvant therapy (35%) were chiefly treated with 5FU or gemcitabine-based regimens, but in this subgroup there was no significant increase in patient survival. The median survival times of patients after surgery from each centre were indistinguishable. Immunohistochemistry was performed on 5 μm sections that were prepared using silanized slides (Menzel Gläser). Staining was performed with the Benchmark System (Ventana), using rabbit anti-USP9X antibody, 1:200 (A301-351A, Bethyl) and anti-ITCH, 1:200 (611198, BD); and the protocol UltraView HRP, with the CC1 modified protocol as pre-treatment. Slides were counterstained with haematoxylin. Staining intensities were scored as absent (0), weak (1), medium (2) and strong (3). For further analysis the staining intensities were grouped as negative (0) and positive (1–3). The Cox regression model assumption of proportional hazard was tested using a plot of the cumulative hazards function.

A second cohort of patient samples was obtained from the Gastrointestinal Cancer Rapid Medical Donation Program in the Department of Pathology at Johns Hopkins Hospital, USA. Use of all human tissue samples from resection specimens and autopsy participants was approved by Johns Hopkins Institutional Review Board, and obtained after informed consent. All samples were collected within 12 h post mortem and formalin fixed before paraffin embedding. Five-micrometre sections were cut from matched primary and metastasis samples onto glass slides. Slides were first incubated in Dako Target Retrieval Solution for antigen retrieval. Slides were then incubated with rabbit anti-USP9X antibody, 1:1,000 (ab26334, Abcam) or 1:200 (NBP1-48321, Novus Biologicals), and anti-SMAD4 as previously described[38]. Signal detection for immunohistochemistry was accomplished with Dako LSAB+System-HRP. Slides were counterstained with haematoxylin.

An additional cohort of pancreatic cancer resection samples was prospectively acquired through the Australian Pancreatic Cancer Network and the Australian Pancreatic Cancer Genome Initiative (http://www.pancreaticcancer.net.au/apgi). Consent was obtained for genomic sequencing through the Australian Pancreatic Cancer Genome Initiative (APGI) for each individual patient following approval from Human Research Ethics Committees (HREC) at participating sites (Sydney South West Area Health Service HREC Western Zone, 2006/054; Sydney Local Health Network HREC RPA Zone, X11-0220; and North Sydney Central Coast Health, Harbour HREC, 0612-251M). We extracted RNA from tumour samples using the Qiagen Allprep kit (Qiagen) in accordance with the manufacturer's instructions, assayed for quality on an Agilent Bioanalyzer 2100 (Agilent Technologies), and subsequently hybridized to Illumina Human HT-12 V4 microarrays. Raw idat files were processed using IlluminaGeneExpressionIdatReader (M. Cowley *et al.*, manuscript in preparation). After array quality control, these data were vs.t transformed, and then robust spline normalized, using the lumi R/Bioconductor package. For the ICGC/APGI cohort, we assumed a proportional hazard: that the probability of death is the same for those censored as for those remaining on study.

For the TMA and expression array cohorts, median survival was estimated using the Kaplan–Meier method and the difference was tested using the log-rank test. P values of less than 0.05 were considered statistically significant. For the TMA cohort, as few parameters were significant in univariate analysis, all were initially considered for Cox Proportional Hazard multivariate analysis in a backward elimination model, and assessed with the SPSS18 Software (IBM) with overall survival used as the primary endpoint. For the ICGC/APGI cohort, clinico-pathological variables analysed with a P value of less than 0.25 on log-rank test were entered into Cox proportional hazard multivariate analysis and the model was resolved using backward elimination. Statistical analysis was performed using StatView 5.0 Software (Abacus Systems). Disease-specific survival was used as the primary endpoint.

27. Tompers, D. M. & Labosky, P. A. Electroporation of murine embryonic stem cells: a step-by-step guide. *Stem Cells* **22,** 243–249 (2004).
28. de Ridder, J., Uren, A., Kool, J., Reinders, M. & Wessels, L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLOS Comput. Biol.* **2,** e166 (2006).
29. Grützmann, R. *et al.* Gene expression profiling of microdissected pancreatic ductal carcinomas using high-density DNA microarrays. *Neoplasia* **6,** 611–622 (2004).
30. Pilarsky, C. *et al.* Activation of Wnt signalling in stroma from pancreatic cancer identified by gene expression profiling. *J. Cell. Mol. Med.* **12,** 2823–2835 (2008).
31. Oberdoerffer, P. *et al.* Efficiency of RNA interference in the mouse hematopoietic system varies between cell types and developmental stages. *Mol. Cell. Biol.* **25,** 3896–3905 (2005).
32. Nathan, J. A. *et al.* The ubiquitin E3 ligase MARCH7 is differentially regulated by the deubiquitylating enzymes USP7 and USP9X. *Traffic* **9,** 1130–1145 (2008).
33. Murray, R. Z., Jolly, L. A. & Wood, S. A. The FAM deubiquitylating enzyme localizes to multiple points of protein trafficking in epithelia, where it associates with E-cadherin and beta-catenin. *Mol. Biol. Cell* **15,** 1591–1599 (2004).
34. Olive, K. P. *et al.* Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science* **324,** 1457–1461 (2009).
35. Ouyang, H. *et al.* Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am. J. Pathol.* **157,** 1623–1631 (2000).
36. Furukawa, T. *et al.* Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16. *Am. J. Pathol.* **148,** 1763–1770 (1996).
37. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30,** 256–268 (2003).
38. Iacobuzio-Donahue, C. A. *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *J. Clin. Oncol.* **27,** 1806–1813 (2009).

doi:10.1038/nature11090

Control of a *Salmonella* virulence locus by an ATP-sensing leader messenger RNA

Eun-Jin Lee[1,2] & Eduardo A. Groisman[1,2]

The facultative intracellular pathogen *Salmonella enterica* resides within a membrane-bound compartment inside macrophages[1]. This compartment must be acidified for *Salmonella* to survive within macrophages[2], possibly because acidic pH promotes expression of *Salmonella* virulence proteins[3,4]. We reasoned that *Salmonella* might sense its surroundings have turned acidic not only upon protonation of the extracytoplasmic domain of a protein sensor[5] but also by an increase in cytosolic ATP levels, because conditions that enhance the proton gradient across the bacterial inner membrane stimulate ATP synthesis[6,7]. Here we report that an increase in cytosolic ATP promotes transcription of the coding region for the virulence gene *mgtC*, which is the most highly induced horizontally acquired gene when *Salmonella* is inside macrophages[8]. This transcript is induced both upon media acidification and by physiological conditions that increase ATP levels independently of acidification. ATP is sensed by the coupling/uncoupling of transcription of the unusually long *mgtC* leader messenger RNA and translation of a short open reading frame located in this region. A mutation in the *mgtC* leader messenger RNA that eliminates the response to ATP hinders *mgtC* expression inside macrophages and attenuates *Salmonella* virulence in mice. Our results define a singular example of an ATP-sensing leader messenger RNA. Moreover, they indicate that pathogens can interpret extracellular cues by the impact they have on cellular metabolites.

The *mgtC* gene is required for survival inside macrophages and for growth in low Mg^{2+} media in several phylogenetically unrelated intracellular pathogens including *Salmonella enterica*[9], *Yersinia pestis*[10], *Brucella suis*[11], *Burkholderia cenocepacia*[12] and *Mycobacterium tuberculosis*[13]. In *Salmonella*, *mgtC* heads the *mgtCBR* operon, which specifies the inner membrane protein MgtC, the Mg^{2+} transporter MgtB and the MgtR peptide promoting MgtC degradation[14]. Transcription from the *mgtCBR* operon is controlled by the virulence regulatory system PhoP/PhoQ[15]. Expression of *mgtC* must be tightly regulated for a normal course of infection because inactivation of the *mgtC* gene attenuates *Salmonella* virulence in mice[9], whereas preventing transcription of AmgR, a PhoP-dependent anti-sense RNA that promotes the preferential degradation of the *mgtC* portion of the polycistronic *mgtCBR* message, renders *Salmonella* hypervirulent[16].

Mild acidic pH induces expression of the *mgtC*[17] (Supplementary Fig. 1) and *mgtB*[18] genes in a *phoP*-dependent manner in wild-type *Salmonella*. This could be ascribed to acidic pH sensing by PhoQ, the cognate sensor of PhoP[5]. However, *Salmonella* seems to use a different mechanism to promote *mgtC* expression in response to mild acidic pH: when bacteria were switched from media at pH 7.7 to media at pH 5.1, the messenger RNA (mRNA) levels of the *mgtC* (Fig. 1a) and *mgtB* (Fig. 1b) coding regions increased even in a *phoP* phoQ*$^-$ strain, which lacks the pH sensor PhoQ and harbours a PhoP allele that can function in its absence[19]. This stimulation seems to be specific to *mgtC* because the mRNA levels corresponding to the *mgtA* coding region did not increase (Fig. 1c), despite the *mgtA* promoter also being under direct transcriptional control of PhoP[15].

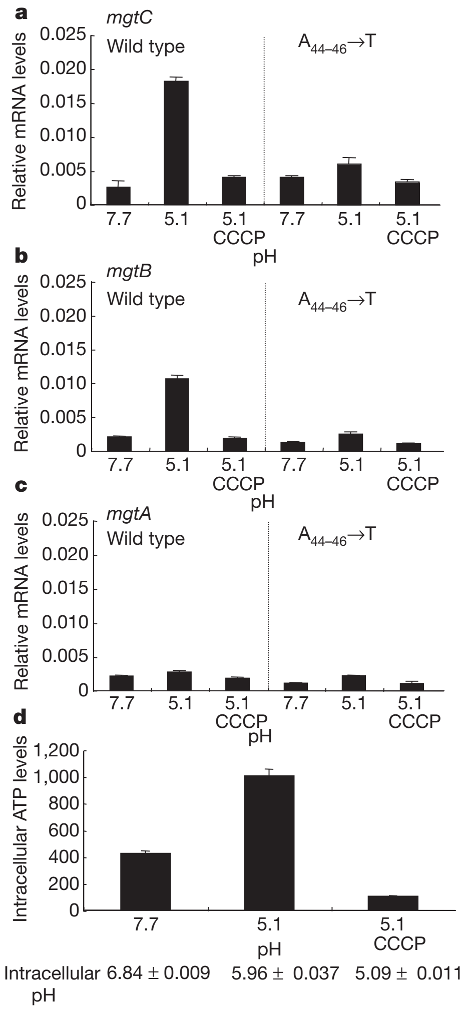


Figure 1 | Mild acidic pH promotes transcription of the *mgtC* and *mgtB* coding regions in a *Salmonella* strain lacking the extracytoplasmic pH sensor PhoQ. Relative mRNA levels of the coding regions of the *mgtC* (**a**), *mgtB* (**b**) and *mgtA* (**c**) genes produced by a *Salmonella phoP* phoQ* strain (EG10232) or a derivative with conserved adenine nucleotides at position 44–46 in the *mgtCBR* leader substituted by thymine nucleotides (EL486). Bacteria were grown in N-minimal media with 500 μM Mg^{2+} at pH 7.7 for 1 h and then for an additional 1 h in the same media at pH 5.1 in the absence or presence of 0.5 μM of the protonophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP). mRNA levels of target genes were normalized to that of 16S ribosomal RNA *rrs* gene. The mean and s.d. from two independent experiments are shown. **d**, Intracellular ATP levels and pH of *Salmonella* (EG10232) grown in media with pH 7.7 and 1 h after being switched to media with pH 5.1 to an attenuance at 600 nm ($D_{600\,nm}$) of 0.453. Values are in picomoles of ATP per millilitre of cells per $D_{600\,nm}$. The mean and s.d. from two independent measurements are shown.

[1]Howard Hughes Medical Institute, Yale School of Medicine, Section of Microbial Pathogenesis, Boyer Center for Molecular Medicine, 295 Congress Avenue, New Haven, Connecticut 06536-0812, USA. [2]Yale Microbial Diversity Institute, PO Box 27389, West Haven, Connecticut 06516, USA.

The stimulation of *mgtC* expression promoted by mild acidic pH could be mediated by an increase in the proton gradient across the inner membrane, hence creating a change in cytosolic ATP. Indeed, the intracellular pH dropped by ~0.9 units and the ATP concentration rose ~2.5-fold within 1 h of switching the *phoP** *phoQ*$^{-}$ strain from media at pH 7.7 to media at pH 5.1 (Fig. 1d). Dissipation of the proton gradient with a protonophore decreased ATP levels (Fig. 1d) and prevented induction of *mgtC* and *mgtB* at pH 5.1 (Fig. 1a, b). (A similar drop in intracellular pH was shown by wild-type *Salmonella*; that is, from pH 8.15 ± 0.112 in media with pH 7.7 to pH 7.25 ± 0.079 in media with pH 5.1 and to pH 5.04 ± 0.031 in media with pH 5.1 in the presence of the protonophore.)

That cellular ATP is the signal upregulating *mgtC* transcription when *Salmonella* experiences mild acidification is supported by two additional experiments performed with strains carrying plasmid pGFP303, which harbours the nucleotide sequence corresponding to the natural *mgtCBR* promoter and leader region fused to a promoterless *gfp* gene at the *mgtC* start codon. Fluorescence was sixfold higher when a purine auxotroph was grown in defined media with high adenine than with low adenine (Fig. 2a). The higher fluorescence is not due to differences in growth rates (data not shown) or a decrease in cellular pH (Fig. 2b), but reflects the higher ATP levels present in bacteria grown at the higher adenine concentration (Fig. 2b). Moreover, it is specific to ATP because a change in uridine levels in the growth media did not alter the fluorescence of a uridine auxotroph harbouring pGFP303 (data not shown). Furthermore, it requires *mgtCBR* regulatory sequences because changes in the adenine concentration in the media did not affect the fluorescence produced by the adenine auxotroph harbouring the plasmid vector (Fig. 2a). And when wild-type *Salmonella* harbouring pGFP303 was grown in glucose, fluorescence was two times higher than when it was grown in glycerol (Fig. 2c), reflecting the larger ATP amounts generated when glucose is used as carbon source (Fig. 2d).

We analysed the *mgtCBR* leader mRNA seeking sequence and structural features that might indicate how it might sense ATP. We identified two short open reading frames (ORFs)—designated *mgtM* and *mgtP*—preceded by putative ribosome-binding sites (Fig. 3a). Here, we focus on *mgtM* because, as subsequently described, it mediates the response to ATP. We determined that *mgtM* is translated *in vivo* (Supplementary Fig. 2) and that similarly sized ORFs preceded by potential ribosome-binding sites are conserved in the *mgtCBR* leader regions from other enteric bacteria (Supplementary Fig. 3). The sequences adjacent to and including *mgtM* have the potential to adopt two alternative structures—stem-loops A and B—in *Salmonella* (Fig. 3a) and other examined species (Supplementary Fig. 3). In-line probing experiments verified the formation of stem-loop B in the wild-type *mgtCBR* leader RNA (Supplementary Fig. 4) and of stem-loop A in a mutant leader RNA with the G_{95}→C substitution, which is predicted to hinder formation of stem-loop B (Supplementary Fig. 4).

The deduced amino-acid sequence of *mgtM* is not conserved in other species (Supplementary Fig. 5), but its length and location relative to stem-loop A, as well as the presence of adenine nucleotides near the base and within stem-loop A, are (that is, positions 44–46 and 56–57; Fig. 3b). Because the left arm of stem-loop A includes the last four *mgtM* codons (Fig. 3a), translation of the complete *mgtM* is anticipated to hinder formation of stem-loop A and favour formation of stem-loop B. Therefore, changes in intracellular ATP levels might affect the

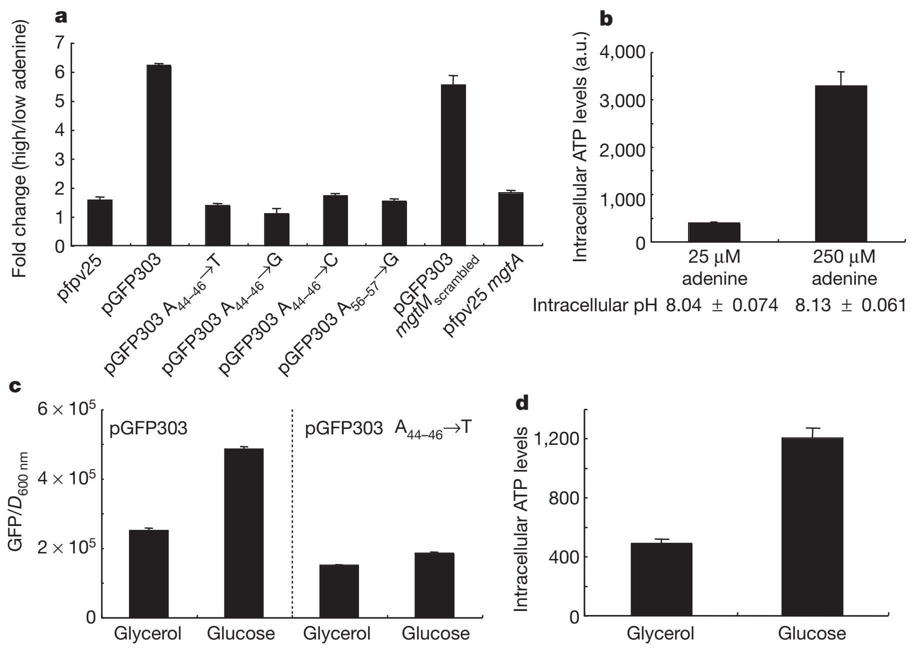


Figure 2 | ATP promotes gene transcription in a manner dependent on conserved adenine nucleotides in the *mgtCBR* leader region. **a**, Fold change in fluorescence produced by an adenine auxotroph (EG9652) harbouring the plasmid vector pfpv25, pGFP303 or derivatives with substitutions of conserved adenine nucleotides at position 44–46 by thymine nucleotides (pGFP303 $A_{44–46}$→T), guanine nucleotides (pGFP303 $A_{44–46}$→G) or cytosine nucleotides (pGFP303 $A_{44–46}$→C), with substitutions of conserved 55–56 adenine nucleotides by guanine nucleotides (pGFP303 $A_{56–57}$→G), or with substitution of most of the *mgtM* sequence except for adenine nucleotides at 44–46 and 55–56 (pGFP303 $mgtM_{scrambled}$). Plasmid pfpv25mgtA harbours a transcriptional fusion between the PhoP-dependent *mgtA* promoter and *mgtA* leader sequence and a promoterless *gfp* gene. Bacteria were grown in N-minimal media with 10 μM Mg^{2+} in the presence of either 25 μM (low) or 250 μM (high) adenine. Fluorescence was monitored as described in Methods. Fold change was calculated by dividing the fluorescence of cells grown in high adenine by that of cells grown in low adenine. Note that the ratio of fluorescence shown at high and low adenine is the same for bacteria harbouring pfpv25mgtA or the vector pfpv25. Data correspond to a representative of four independent experiments conducted in duplicate. **b**, Intracellular ATP levels and intracellular pH of *purB* *Salmonella* (EG9652) grown in the presence of the indicated concentrations of adenine were determined by labelling with phosphorus-32 and then separating by thin-layer chromatography. **c**, Fluorescence produced by wild-type *Salmonella* (14028s) harbouring pGFP303 or pGFP303 $A_{44–46}$→T after growth for 4 h in N-minimal media with 10 μM Mg^{2+} in the presence of either 38 mM glycerol (~0.35%) or 0.2% glucose as a carbon source. GFP, green fluorescent protein. **d**, Intracellular ATP levels of cells grown in the indicated carbon sources to $D_{600\,nm}$ = 0.419 were determined by luminescence as described in Methods. The mean and s.d. from two independent measurements are shown.

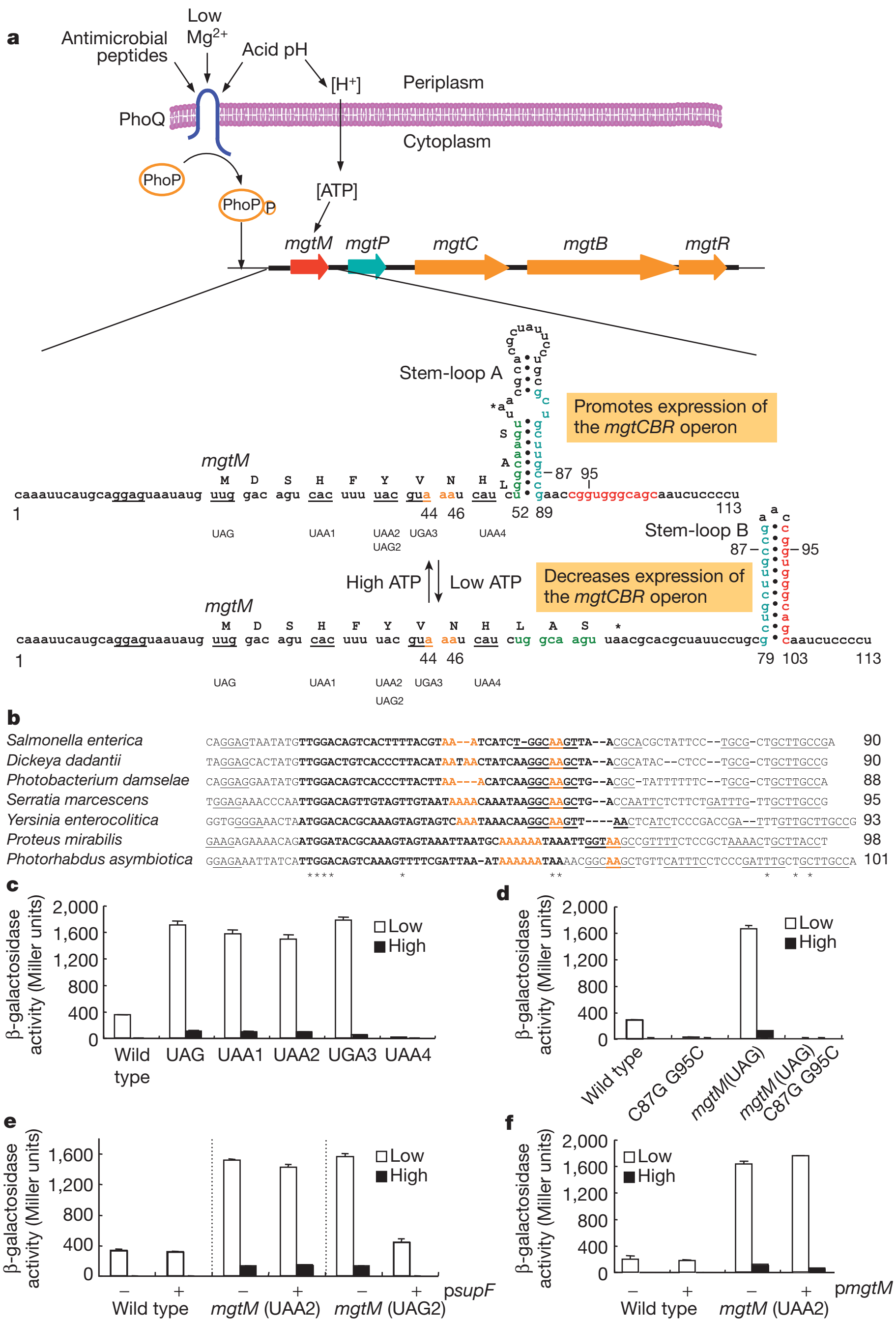


Figure 3 | Regulation of the *Salmonella mgtCBR* virulence operon by the PhoP/PhoQ system and *mgtCBR* leader region. a, The sensor PhoQ responds to extracytoplasmic low Mg^{2+}, acidic pH and antimicrobial peptides by promoting phosphorylation of the PhoP protein, which binds to the *mgtCBR* promoter, resulting in transcription initiation. Acidic pH also produces a proton gradient across the inner membrane, resulting in higher intracellular ATP levels. Intracellular ATP levels control transcription elongation into the *mgtCBR* coding regions by affecting the coupling/uncoupling of transcription of the *mgtCBR* leader and translation of the short ORF *mgtM*, which affects the formation of alternative stem-loops A or B. The *mgtM* ribosome-binding site is underlined. Positions and sequences of stop codon mutations or nucleotide substitutions in the strains used in the experiments presented in **c** and **d** are indicated and denoted below the *mgtM* sequence. **b**, Conservation of adenine nucleotides in *mgtM* relative to stem-loop structures in the *mgtC* leader. Alignment of the nucleotide sequences corresponding to *mgtM* from the indicated species. Sequences in bold correspond to *mgtM*. The *mgtM* ribosome-binding sites are underlined. Asterisks correspond to nucleotides conserved in all listed species. Conserved adenine nucleotides are coloured in orange; sequences with a potential to adopt stem-loop structures are underlined. **c**, β-galactosidase activity (Miller units) produced by *Salmonella* with a chromosomal *mgtC–lac* fusion (EG9527) and isogenic derivatives with mutation of the start codon (EG19269) or with stop codons at different positions (EG19285, EG19289, EG19293 and EG19298) in *mgtM*. Bacteria were grown in N-minimal media containing low (10 μM) or high (10 mM) Mg^{2+} for 4 h. **d**, β-galactosidase activity (Miller units) produced by *Salmonella* with a chromosomal *mgtC–lac* fusion (EL92) and isogenic derivatives with $C_{87}{\rightarrow}G$ and $G_{95}{\rightarrow}C$ substitution (EL96) or mutation of the start codon in *mgtM* (EL97) or both (EL98). Bacteria were grown as described in **c**. The mean and s.d. from three independent experiments are shown in **c** and **d**. **e**, β-galactosidase activity produced by *Salmonella* with a chromosomal *mgtC–lac* fusion harbouring either plasmid psupF or the empty vector pUH21-2lacI[q] (EL86, EL87) and isogenic derivatives with an amber stop codon (EL90, EL91) or ochre stop codon (EL88, EL89) at position 39–41 of *mgtM*. Bacteria were grown as described except in the presence of ampicillin and 0.2 mM IPTG. **f**, β-galactosidase activity produced by *Salmonella* with a chromosomal *mgtC–lac* fusion (EG9527) or an isogenic strain with a stop codon at position 39–41 in *mgtM* (EG19289) harbouring either the vector or plasmid pmgtM. Bacteria were grown as described except in the presence of ampicillin and 0.1 mM IPTG. The mean and s.d. from two independent experiments are shown in **e** and **f**.

coupling/uncoupling of transcription of the *mgtCBR* leader with translation of *mgtM*. This would determine whether stem-loop A or B forms, thereby dictating whether transcription continues into the *mgtCBR* coding regions. This is analogous to the mechanism by which cytoplasmic UTP levels control expression of the pyrimidine biosynthetic gene *pyrBI* in *Escherichia coli*, except that low UTP promotes *pyrBI* transcription when RNAP pauses at a U-rich segment of the *pyrBI* leader mRNA[20] whereas we propose that high ATP furthers expression of the *mgtCBR* coding regions.

In agreement with the notion that a transcription attenuation-like mechanism[20,21] regulates transcription elongation into the *mgtCBR* coding region, a strain with a chromosomal mutation in the *mgtM* start codon and harbouring a *lac* fusion in the *mgtC* coding region produced over four times more β-galactosidase than the isogenic strain with the wild-type *mgtC* leader when grown in low Mg^{2+} to induce the PhoP/PhoQ system[15] (Fig. 3c, d). Likewise, strains harbouring stop codons at the fourth, sixth or seventh positions of *mgtM* (Fig. 3a) also derepressed *mgtC–lac* expression (Fig. 3c). This derepression is due to a defect in *mgtM* translation (as opposed to resulting from an effect on the structure of the *mgtC* leader mRNA): a plasmid expressing the amber suppressor *supF* restored normal *mgtC–lac* transcription to an *mgtM* mutant with an amber stop codon at the sixth position but not to an isogenic mutant with an ochre stop codon at the same position (Fig. 3e). As expected, the *supF*-expressing plasmid had no effect on *mgtC–lac* transcription in a strain harbouring the wild-type *mgtC* leader (Fig. 3e). An *mgtM* derivative with a stop codon at the ninth position failed to express *mgtC–lac* (Fig. 3c), possibly because translation of *mgtM* beyond the eighth codon would result in formation of stem-loop B (Fig. 3a). Translation of *mgtM* exerts its regulatory effect on the associated *mgtCBR* coding regions in *cis* (as opposed to *mgtM* encoding a *trans*-acting peptide) because a plasmid expressing the *mgtM* ORF failed to restore normal *mgC–lac* expression to an *mgtM* stop codon mutant, behaving like the vector control (Fig. 3f); and it had no effect on a strain with a wild-type *mgtC* leader (Fig. 3f). Our data indicate that whenever translation stops before the ribosome reaches the ninth *mgtM* codon, stem-loop A forms, which promotes expression of the *mgtCBR* coding region; and when *mgtM* is translated beyond a certain position, stem-loop B is favoured, which reduces transcription of the *mgtCBR* coding region (Fig. 3a). As predicted, the pronounced *mgtC–lac* derepression shown by the *mgtM* start codon mutant (Fig. 3c, d) was eliminated by the simultaneous introduction of $C_{87}{\rightarrow}G$ and $G_{95}{\rightarrow}C$ substitutions (Fig. 3d), which favours formation of stem-loop B (Fig. 3a).

The phenotypes described above are not an artefact resulting from the absence of a functional *mgtC* gene in the *mgtC–lac* strains. This is because mutation of the *mgtM* start codon derepressed *mgtC* levels in an otherwise wild-type *Salmonella* that experienced mild acidification (Supplementary Fig. 6), and because introduction of a stop codon at the ninth position of *mgtM* silenced *mgtC* expression in an isogenic strain subjected to the same conditions (Supplementary Fig. 6).

The conserved adenine nucleotides at positions 44–46 or 56–57 of the *mgtCBR* leader (Fig. 3a) are critical for the response to ATP: on the one hand, the mild acid induction of the *mgtC* and *mgtB* genes disappeared in a strain with a chromosomal mutation at position 44–46 in the *mgtCBR* leader (Fig. 1a, b). Likewise, growth in glucose no longer promoted higher fluorescence than growth in glycerol in wild-type *Salmonella* carrying a pGFP303 derivative with the adenine nucleotides at position 44–46 substituted by thymine nucleotides (Fig. 2c). Yet, this mutant leader retained a wild-type structure (Supplementary Fig. 7) and ability to respond to Mg^{2+} (Supplementary Fig. 8)[22]. Moreover, changes in the adenine concentration in the media failed to affect the fluorescence produced by the purine auxotroph with derivatives of pGFP303 where the adenine nucleotides at position 44–46 were replaced by thymine nucleotides, guanine nucleotides or cytosine nucleotides, or with a plasmid with the adenine nucleotides at position 56–57 replaced by guanine nucleotides (to avoid destabilizing stem-loop A, the last two nucleotides were not replaced by thymine nucleotides or cytosine nucleotides; Fig. 3a) (Fig. 2a). On the other hand, an engineered *mgtCBR* leader with a scrambled *mgtM* sequence that retained the adenine nucleotides at positions 44–46 and 56–57 had a wild-type response to ATP (Fig. 2a). The RNAs from the wild-type and ATP-blind mutant *mgtC* leader had indistinguishable *in vitro* profiles when investigated at two pHs (Supplementary Fig. 7b). This provides further support to the notion that the increase in *mgtCBR* expression resulting from mild acidification is mediated by cytosolic ATP levels as opposed to protons being sensed directly by the *mgtC* leader mRNA. Cumulatively, these data provide a singular example of a bacterial mRNA leader that senses ATP, using a mechanism that is different from those previously described in eukaryotic organisms[23–25].

We wondered whether ATP sensing by the *mgtCBR* mRNA leader is required for *Salmonella* virulence given that *mgtC* is expressed in a variety of tissues during infection in several different animal hosts[8,26–28] and that inactivation of the *mgtC* gene hinders survival inside macrophages and virulence in mice[9]. We established that when *Salmonella* was inside the macrophage-like cell line J774.A1, the mRNA levels of the *mgtA* and *mgtC* leaders region rose in parallel (Supplementary Fig. 9), which probably reflects activation of their respective promoters during

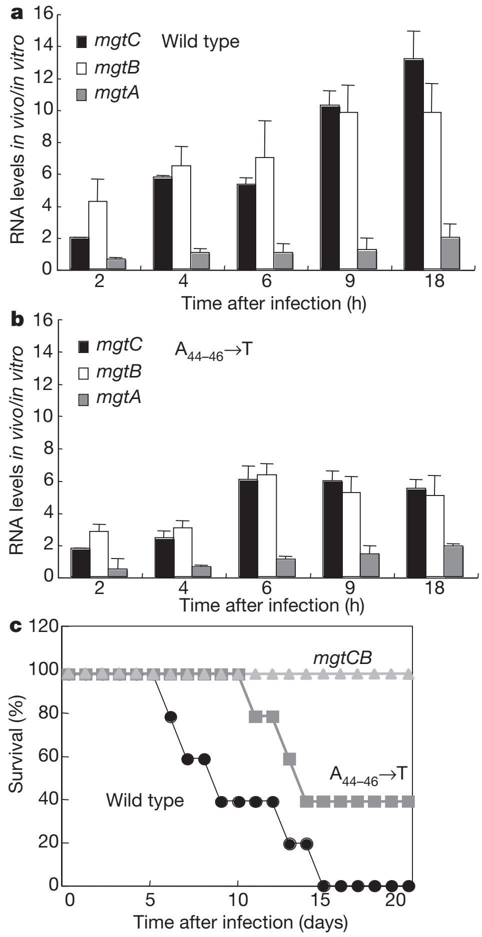


Figure 4 | Expression of the *mgtC* coding region inside macrophages is dependent on its leader region's ability to sense ATP, a property required for *Salmonella* virulence. **a**, Relative mRNA levels of the *mgtC*, *mgtB* and *mgtA* coding regions of wild-type *Salmonella* harbouring the wild-type *mgtCBR* leader (14028s) inside J774 A.1 macrophages at the indicated times after infection. **b**, The same as in **a** except using a mutant *Salmonella* with the chromosomal adenine nucleotides at position 44–46 of the *mgtCBR* leader substituted by thymine nucleotides (EL341). **c**, Survival of C3H/HeN mice inoculated intraperitoneally with $\sim 10^4$ colony-forming units of wild-type *Salmonella* (14028s) or ATP-sensing defective mutant (EL341) or deleted for both the *mgtC* and *mgtB* coding regions (EL6).

infection[29]. By contrast, the mRNA levels corresponding to the *mgtC* and *mgtB* coding regions increased markedly (that is, tenfold by 9 h after internalization relative to *Salmonella* grown in tissue culture media; Fig. 4a), but no induction was observed for the coding region of the *mgtA* gene (Fig. 4a), which is not required for *Salmonella* virulence[9]. This might reflect differences between the *mgtCBR* (Fig. 3a) and *mgtA*[30] leaders. Indeed, the mRNA levels of the *mgtA* coding region did not increase under conditions promoting higher ATP levels (Fig. 2a) or mild acidification (Fig. 1c). Induction of the *mgtC* and *mgtB* coding regions inside macrophages requires the ability of the *mgtCBR* leader mRNA to sense ATP because it was defective in a chromosomal mutant with the adenine nucleotides at position 44–46 replaced by thymine nucleotides (Fig. 4b). This mutant was attenuated for virulence after intraperitoneal inoculation of mice (Fig. 4c), albeit not as much as a strain deleted for the *mgtC* and *mgtB* coding regions (Fig. 4c), implying that other signalling inputs remain functional in this mutant.

Cumulatively, our findings demonstrate that the *mgtCBR* leader senses cytosolic ATP levels, thereby enabling differential control of the *mgtCBR* operon from that of other PhoP-activated genes. This property is critical for *Salmonella* virulence, possibly because *Salmonella* resides within a phagosome that is mildly acidic[1], which is a condition that can generate high ATP levels in the bacterium[6,7]. Finally, our data highlight how pathogens can interpret host-derived signals, such as acidic pH, by the changes they experience in their cellular constituents.

METHODS SUMMARY

Bacterial strains and plasmids used in this study are listed in Supplementary Table 1. All *S. enterica* serovar Typhimurium strains were derived from the wild-type strain 14028s and were constructed by phage P22-mediated transductions as described. All DNA oligonucleotides are listed in Supplementary Table 2. Detailed descriptions of examining the effect of pH or ATP on gene expression, determining intracellular ATP levels, examining expression inside macrophage and mouse virulence assays are in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 December 2011; accepted 26 March 2012.

1. Garcia-del Portillo, F. *Salmonella* intracellular proliferation: where, when and how? *Microbes Infect.* **3,** 1305–1311 (2001).
2. Rathman, M., Sjaastad, M. D. & Falkow, S. Acidification of phagosomes containing *Salmonella typhimurium* in murine macrophages. *Infect. Immun.* **64,** 2765–2773 (1996).
3. Alpuche Aranda, C. M., Swanson, J. A., Loomis, W. P. & Miller, S. I. *Salmonella typhimurium* activates virulence gene transcription within acidified macrophage phagosomes. *Proc. Natl Acad. Sci. USA* **89,** 10079–10083 (1992).
4. Yu, X. J., McGourty, K., Liu, M., Unsworth, K. E. & Holden, D. W. pH sensing by intracellular *Salmonella* induces effector translocation. *Science* **328,** 1040–1043 (2010).
5. Prost, L. R. *et al.* Activation of the bacterial sensor kinase PhoQ by acidic pH. *Mol. Cell* **26,** 165–174 (2007).
6. Harold, F. M. & Maloney, P. C. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhart, F. C. *et al.*) 283–306 (American Society for Microbiology, 1996).
7. Senior, A. E. The proton-translocating ATPase of *Escherichia coli. Annu. Rev. Biophys. Biophys. Chem.* **19,** 7–41 (1990).
8. Eriksson, S., Lucchini, S., Thompson, A., Rhen, M. & Hinton, J. C. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica. Mol. Microbiol.* **47,** 103–118 (2003).
9. Blanc-Potard, A. B. & Groisman, E. A. The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. *EMBO J.* **16,** 5376–5385 (1997).
10. Grabenstein, J. P., Fukuto, H. S., Palmer, L. E. & Bliska, J. B. Characterization of phagosome trafficking and identification of PhoP-regulated genes important for survival of *Yersinia pestis* in macrophages. *Infect. Immun.* **74,** 3727–3741 (2006).
11. Lavigne, J. P., O'Callaghan, D. & Blanc-Potard, A. B. Requirement of MgtC for *Brucella suis* intramacrophage growth: a potential mechanism shared by *Salmonella enterica* and *Mycobacterium tuberculosis* for adaptation to a low-Mg^{2+} environment. *Infect. Immun.* **73,** 3160–3163 (2005).
12. Maloney, K. E. & Valvano, M. A. The *mgtC* gene of *Burkholderia cenocepacia* is required for growth under magnesium limitation conditions and intracellular survival in macrophages. *Infect. Immun.* **74,** 5477–5486 (2006).
13. Buchmeier, N. *et al.* A parallel intraphagosomal survival strategy shared by *Mycobacterium tuberculosis* and *Salmonella enterica. Mol. Microbiol.* **35,** 1375–1382 (2000).
14. Alix, E. & Blanc-Potard, A. B. MgtC: a key player in intramacrophage survival. *Trends Microbiol.* **15,** 252–256 (2007).
15. Soncini, F. C., Garcia Vescovi, E., Solomon, F. & Groisman, E. A. Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. *J. Bacteriol.* **178,** 5092–5099 (1996).
16. Lee, E. J. & Groisman, E. A. An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol. Microbiol.* **76,** 1020–1033 (2010).
17. Retamal, P., Castillo-Ruiz, M. & Mora, G. C. Characterization of MgtC, a virulence factor of *Salmonella enterica* serovar Typhi. *PLoS ONE* **4,** e5551 (2009).
18. Bearson, B. L., Wilson, L. & Foster, J. W. A low pH-inducible, PhoPQ-dependent acid tolerance response protects *Salmonella typhimurium* against inorganic acid stress. *J. Bacteriol.* **180,** 2409–2417 (1998).
19. Chamnongpol, S. & Groisman, E. A. Acetyl phosphate-dependent activation of a mutant PhoP response regulator that functions independently of its cognate sensor kinase. *J. Mol. Biol.* **300,** 291–305 (2000).
20. Turnbough, C. L. Jr & Switzer, R. L. Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol. Mol. Biol. Rev.* **72,** 266–300 (2008).
21. Merino, E. & Yanofsky, C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21,** 260–264 (2005).
22. Spinelli, S. V., Pontel, L. B., Garcia Vescovi, E. & Soncini, F. C. Regulation of magnesium homeostasis in *Salmonella*: Mg^{2+} targets the *mgtA* transcript for degradation by RNase E. *FEMS Microbiol. Lett.* **280,** 226–234 (2008).
23. Amiott, E. A. & Jaehning, J. A. Mitochondrial transcription is regulated via an ATP "sensing" mechanism that couples RNA abundance to respiration. *Mol. Cell* **22,** 329–338 (2006).
24. Dennis, P. B. *et al.* Mammalian TOR: a homeostatic ATP sensor. *Science* **294,** 1102–1105 (2001).
25. Shu, D. & Guo, P. A viral RNA that binds ATP and contains a motif similar to an ATP-binding aptamer from SELEX. *J. Biol. Chem.* **278,** 7119–7125 (2003).
26. Harvey, P. C. *et al. Salmonella enterica* serovar Typhimurium colonizing the lumen of the chicken intestine are growing slowly and up-regulate a unique set of virulence and metabolism genes. *Infect. Immun.* **79,** 4105–4121 (2011).
27. Lawley, T. D. *et al.* Genome-wide screen for *Salmonella* genes required for long-term systemic infection of the mouse. *PLoS Pathog.* **2,** e11 (2006).
28. Rehfuss, M. Y., Parker, C. T. & Brandl, M. T. *Salmonella* transcriptional signature in *Tetrahymena* phagosomes and role of acid tolerance in passage through the protist. *ISME J.* **5,** 262–273 (2011).
29. Heithoff, D. M. *et al.* Bacterial infection as assessed by *in vivo* gene expression. *Proc. Natl Acad. Sci. USA* **94,** 934–939 (1997).
30. Park, S. Y., Cromie, M. J., Lee, E. J. & Groisman, E. A. A bacterial mRNA leader that employs different mechanisms to sense disparate intracellular signals. *Cell* **142,** 737–748 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Turnbough for discussions, M. Wade for help with the mouse virulence assays, and R. Breaker and A. Roth for help with in-line probing experiments. This work was supported, in part, by grant AI49561 from the National Institutes of Health to E.A.G., who is an investigator of the Howard Hughes Medical Institute.

Author Contributions E.-J.L. conducted the experiments. E.-J.L. and E.A.G. designed the study and wrote the paper. Both authors read the paper and contributed to its final form.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.A.G. (eduardo.groisman@yale.edu).

METHODS

Bacterial strains, plasmids, oligodeoxynucleotides and growth conditions. Bacterial strains and plasmids used in this study are listed in Supplementary Table 1. All *S. enterica* serovar Typhimurium strains were derived from the wild-type strain 14028s[31] and constructed by phage P22-mediated transductions as described[32]. All DNA oligonucletides are listed in Supplementary Table 2. Bacteria were grown at 37 °C in Luria-Bertani broth, N-minimal media (pH 7.7)[33] supplemented with 0.1% casamino acids, 38 mM glycerol and the indicated concentrations of $MgCl_2$. To examine the effect of pH or carbon source on gene expression, we used a modified N-minimal medium containing 0.2% glucose instead of 38 mM glycerol. *E. coli* DH5α was used as the host for preparation of plasmid DNA. Ampicillin was used at 50 μg ml^{-1}, chloramphenicol at 20 μg ml^{-1}, kanamycin at 20 μg ml^{-1}, tetracycline at 10 μg ml^{-1} and fusaric acid at 12 μg ml^{-1}.

Effect of pH on gene expression. Cells were grown overnight in N-minimal medium containing 10 mM Mg^{2+}. Dilution (1/100) of the overnight culture was used to inoculate 20 ml of the same medium and grown for 3 h. Cells were then washed and transferred to 20 ml of N-minimal medium containing 500 μM Mg^{2+} and grown for 1 h. The cells were collected and washed with N-minimal medium containing 500 μM Mg^{2+} and an aliquot corresponding to 1/10 volume of cells for pH 7.7 before switching to pH 5.1. Then, cells were resuspended in 20 ml of N-minimal medium containing 500 μM Mg^{2+} at pH 5.1 with or without 0.5 μM carbonyl cyanide 3-chlorophenylhydrazone and growth continued for 1 h. Bacteria were stabilized using RNAprotect Bacteria Reagent (Qiagen) and RNA was isolated for further analysis.

Effect of exogenous adenine on gene expression. The adenine auxotrophic strain EG9652 harbouring plasmid pGFP303, its derivatives, pfpv25mgtA or the plasmid vector were grown overnight in N-minimal medium containing 10 mM Mg^{2+}, 500 μM adenine and ampicillin. One millilitre of the overnight culture was washed in the N-minimal medium without Mg^{2+} and adenine and resuspended in 1 ml of the same media. The suspended bacteria were inoculated 1/50 volume in N-minimal medium with 10 μM Mg^{2+} and ampicillin in the presence of either 25 or 250 μM adenine. Fluorescence and $D_{600\,nm}$ of the cultures were monitored for 6.5 h with shaking 37 °C in a Victor[3] plate reader (Perkin Elmer). To prevent evaporation, the 96-well plate was covered with mineral oil. GFP expression of a given strain was determined by plotting fluorescence over $D_{600\,nm}$. Note that the *purB* strains were freshly transduced before each experiment to prevent the isolation of suppressor mutations.

Measurement of intracellular ATP in *Salmonella*. We measured intracellular ATP levels using a luminometer (BioTek Synergy H1) as described with modification[34]. Briefly, bacteria were grown overnight in N-minimal media containing 10 mM Mg^{2+} and 0.2% glucose as a carbon source. One millilitre of the overnight culture was washed in the N-minimal medium without Mg^{2+} and resuspended in 1 ml of the same media. Diluted (1/100) bacteria were inoculated in 2 ml of N-minimal media containing 10 μM Mg^{2+} (with proper antibiotics if necessary) and grown for 4 h. Cells were normalized by $D_{600\,nm}$ and resuspended in 500 μl of phosphate-buffered saline (PBS). Nucleic acids were extracted by adding 100 μl of ice-cold 3 M perchloric acid. After incubating for 5 min, extracts were neutralized with 225 μl of neutralization buffer (1 M KOH, 0.5 M Tris, 0.5 M KCl) and centrifuged for 10 min. Fifty microlitres of the supernatant were diluted with 50 μl of L buffer (25 mM KCl, 50 mM $MgSO_4$, 100 mM HEPES pH 7.4) and intracellular ATP was measured using an ATP Determination Kit (Invitrogen) according to the manufacturer's instruction. Intracellular ATP levels (picomoles per millilitre of cells at given $D_{600\,nm}$) were converted using reference to standards of known concentration.

To measure intracellular ATP levels in response to exogenous adenine levels (Fig. 2b), *purB Salmonella* were grown overnight in N-minimal medium containing 10 mM Mg^{2+} and antibiotic. One millilitre of the overnight culture was washed in the N-minimal medium without Mg^{2+} and adenine and resuspended in 1 ml of the same media. Diluted (1/50) suspended bacteria were inoculated in modified N-minimal media containing 10 μM Mg^{2+} and 0.3 mM inorganic phosphate and grown for 6.5 h. Then, 120 μCi of phosphorus-32 were added and labelled for 15 min. The labelled nucleoside 5′-triphosphates were extracted by formic acid extraction as described previously and analysed by PEI-cellulose thin-layer chromatography plate[35]. The levels of intracellular ATP were quantified by a Typhoon FLA-9000 phosphorimager (GE Healthcare).

Examining gene expression inside macrophages. Macrophage infection was performed as described[8] with the following modifications: J774 A.1 macrophages were seeded in six-well plates in RPMI medium (Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum, 2 mM L-glutamine and 10 mM HEPES at a density of 10^6 per well one day before infection with *Salmonella*. Bacteria were grown overnight in Luria-Bertani broth media at 37 °C, washed with PBS and used to infect macrophages at a multiplicity of infection of 100:1. Plates were centrifugated at 1,000*g* for 5 min (defined as time 0) and incubated for 1 h for phagocytosis. Extracellular bacteria were killed with RPMI media containing 50 μg ml^{-1} gentamicin. The media were replaced after 1 h with the same media containing 10 μg ml^{-1} gentamicin and incubation continued as indicated in the legend to Fig. 4a, b. At each time point, infected macrophages were lysed and stabilized with Tri reagent (Applied Biosystems) and RNA was extracted according to the manufacturer's instruction. Control RNA from *Salmonella* cultured in tissue culture media was obtained as described previously[8]. *Salmonella* mRNA levels at each time point were normalized by the mRNA levels after growth in tissue culture media. To estimate the intracellular bacteria at each time point, cells were lysed with PBS containing 0.1% Triton X-100 and plated on Luria-Bertani broth plate with proper dilutions, or DNA was extracted from the same RNA sample according to the manufacturer's instructions (Applied Biosystems) and measured by quantitative real time PCR.

Mouse virulence assays. Six- to eight-week-old female C3H/HeN mice were inoculated intraperitoneally with ~10^4 colony-forming units. Mouse survival was followed for 21 days. Virulence assays were conducted three times with similar outcomes and data correspond to groups of five mice. All animals were housed in temperature- and humidity-controlled rooms, maintained on a 12 h light/12 h dark cycle. All procedures complied with regulations of the Institutional Animal Care and Use Committee of the Yale School of Medicine.

Construction of plasmids harbouring *lacZ* fusions to *mgtM*. PCR fragments corresponding to nucleotides 1–59 of the *mgtCBR* leader were amplified with primer 9801, which includes the sequence corresponding to the $plac_{1-6}$ promoter, and either primer 9802 or 9803 (creating a stop codon) using 14028s genomic DNA as a template. The resulting PCR products were digested with SmaI and XbaI and cloned into plasmid pACYC-′lacZ digested with the same enzymes. The sequences of the resulting constructs were verified by DNA sequencing.

Construction of plasmids harbouring fusions to a promoterless *gfp* gene. pGFP303, a plasmid with the PhoP-dependent *mgtCBR* promoter and the wild-type *mgtC* leader fused to a promoterless *gfp* gene, was constructed as follows: a PCR fragment generated with primers 1746 and 8117 using 14028s genomic DNA as a template and digested with EcoRI and XbaI was cloned into plasmid pfpv25 digested with the same enzymes.

Derivatives of pGFP303 with nucleotide substitution in the *mgtCBR* leader region were constructed by cloning PCR fragments generated by two rounds of PCR reactions. For the A_{44-46}→T substitution in the *mgtC* leader, a first PCR fragment was generated with primers 1746 and 11727, and a second fragment was generated with primers 11726 and 8117 using 14028s genomic DNA as a template. A third PCR was performed with primers 1746 and 8117 using the two PCR-generated DNA fragments as templates. The resulting PCR product was cloned into pfpv25 using the same restriction enzymes used for construction of pGFP303. All other substitutions were generated in a similar way using the following primer pairs: A_{44-46}→G (1746/11725 and 11724/8117), A_{44-46}→C (1746/11399 and 11398/8117), A_{56-57}→G (1746/11723 and 11722/8117) and $mgtM_{scrambled}$ (1746/11950 and 11949/8117).

pfpv25mgtA, a plasmid with the PhoP-dependent *mgtA* promoter and wild-type *mgtA* leader fused to a promoterless *gfp* gene, was constructed as follows: a PCR fragment was generated with primers 6208 and 11737 using 14028s genomic DNA as a template and digested with EcoRI and XbaI and then cloned into pfpv25 digested with the same enzymes. The sequences of the resulting constructs were verified by DNA sequencing.

Construction of a plasmid harbouring the *mgtM* ORF. Plasmid pmgtM was constructed as follows: a PCR fragment corresponding to the *mgtM* gene was generated by PCR with primers 10074 and 10075 using 14028s genomic DNA as a template, digested with HindIII and BamHI and cloned into pUHE 21-2lacI^q digested with the same enzymes.

Construction of a strain with chromosomal deletion of the *mgtC* and *mgtB* genes. A *Salmonella* strain deleted for the *mgtC* and *mgtB* genes was generated by the one-step gene inactivation method[36]. A Cm^R cassette was PCR amplified from plasmid pKD3 using primers 1908 and 1911 for the deletion of the *mgtC* and *mgtB* genes; the resulting PCR products were integrated into 14028s chromosome to generate strain EG16736 (*mgtCB*::Cm^R). The *mgtCB* deletion (EL6) strain was made by removing the antibiotic resistance cassette from EG16736 using plasmid pCP20 as described[36].

Construction of strains with chromosomal mutations in the *mgtCBR* leader region. Two different methods were used to generate strains with chromosomal mutations in the *mgtCBR* leader. To create stop codon mutants in *mgtM* and with nucleotide substitution in the *mgtCBR* leader region, we used the fusaric acid method as described[16]. We introduced a Tet^R cassette in the *mgtCBR* leader region as follows: we generated a 1990-base-pair PCR product harbouring the *tetRA* genes using as template chromosomal DNA from strain MS7953s and primers 7310/7370. The product was purified and used to electroporate 14028s or EG9527 *Salmonella* containing plasmid pKD46. The resulting *mgtCBR leader::tetRA* strains containing plasmid pKD46 (EG18715 and EG18798) were kept at 30 °C.

Then, we replaced the *tetRA* cassette in the *mgtCBR* leader region using DNA fragments carrying a mutation to create nucleotide substitutions. This was accomplished by preparing a DNA fragment harbouring nucleotide substitutions in the *mgtCBR* leader region by a two-step PCR process.

To create strains with stop codon mutations in *mgtM*, DNA fragments carrying stop codons in *mgtM* were prepared as follows: we used two primer pairs 8118/8769 and 8770/7308 (for UAG), 8118/8878 and 8880/7308 (for UAA1), 8118/8881 and 8879/7308 (for UAA2), 8118/8986 and 8987/7308 (for UGA3), 8118/8882 and 8883/7308 (for UAA4) or 8118/9851 and 9850/7308 (for UAG2) and 14028s genomic DNA as a template in the first PCR reaction. For the second PCR reaction, we mixed the two PCR products from the first PCR reaction as templates and amplified the DNA fragment with expected substitution using primers 8118 and 7308. The resulting PCR products were purified and integrated into the EG18798 or EG18715 (for EG19307 and EG19322) chromosome and selected against Tet^R with media containing fusaric acid to generate strains EG19269, EG19285, EG19289, EG19293, EG19298, EL90 (or EL91), EG19307 and EG19322, Tet^S Amp^S chromosomal mutants, respectively.

DNA fragments carrying the A44–46 to T substitution in the *mgtC* leader were prepared by a two-step PCR reaction. For the first PCR reaction, we used two primer pairs 8118/11727 and 11726/7308, and 14028s genomic DNA as a template. For the second PCR reaction, we mixed the two PCR products from the first PCR reaction as templates and amplified the DNA fragment with expected substitution using primers 8118 and 7308. The resulting PCR products were purified and integrated into the EG18715 chromosome and selected against Tet^R with media containing fusaric acid to generate EL341, a Tet^S Amp^S chromosomal mutant. The presence of the expected substitution was verified by sequencing.

All other chromosomal mutants with substitutions in the *mgtCBR* leader were constructed by a multiple-step PCR process. Strain EL92 was constructed by inserting a Cm^R cassette in the *yicL* gene, which is 278 nucleotides upstream from *mgtC* transcription start site. The Cm^R cassette was amplified from plasmid pKD3 using primers 4801 and 4802 and the resulting PCR products were integrated into EG9527 chromosome to generate EL92 (*yicL*::Cm^R). Then, we prepared DNA fragments containing a Cm^R cassette and the proper nucleotide substitutions in the *mgtCBR* leader using two primer pairs and EL92 genomic DNA as a template: primer pairs 10077/8265 and 8264/7308 for the C87→G G95→C substitution, and primer 10077/8769 and 8770/7308 for UAG mutation in the *mgtM*. The two resulting DNA fragments from the first PCR reactions were mixed and used as PCR templates to amplify DNA fragments containing Cm^R cassette and the proper nucleotide substitution using 10077 and 7308. For EL98, DNA fragments were generated using primer pairs 10077/8265 and 8264/7308 and EL97 genomic DNA as a template. The resulting DNA fragments were purified and integrated into EG9527 chromosome by the one-step inactivation method[36] and mutants were selected for resistance to chloramphenicol. The presence of the expected substitution was verified by DNA sequencing.

Quantitative RT–PCR. Total RNA was isolated using RNeasy Kit (Qiagen) according to the manufacturer's instructions. The purified RNA was quantified using a Nanodrop machine (NanoDrop Technologies). Complementary DNA (cDNA) was synthesized using High Capacity RNA-to cDNA Master Mix (Applied Biosystems). The mRNA levels of the *mgtC*, *mgtB*, *mgtA* and *rrs* genes were measured by quantification of cDNA using Fast SYBR Green PCR Master Mix (Applied Biosystems) and appropriate primers (*mgtC* leader: 6962/6963; *mgtC* coding: 7530/7531; *mgtB* coding: 7763/7764; *mgtA* leader: 7225/7226; *mgtA* coding: 4308/4309) and monitored using a Fast ABI7500 machine (Applied Biosystems). Data were normalized to the levels of 16S ribosomal RNA amplified with primers 6970 and 6971.

β-galactosidase assays. Cells were grown overnight in N-minimal media and washed twice in N-minimal media before resuspending them in N-minimal media with different $MgCl_2$ concentrations for 4 h at 37 °C with shaking. The activity was determined as described[37]. Data correspond to two independent experiments conducted in duplicate.

Determining intracellular pH in *Salmonella*. We measured intracellular pH using green fluorescent protein as described with modification[38]. Bacteria harbouring a plasmid containing the *gfp* gene expressed from heterologous promoter (pfpv25.1) were grown as shown in Figs 1 and 2b. Cells were normalized by $D_{600\,nm}$ and resuspended in 150 µl of PBS in a 96-well black microplate. Excitation spectra were measured at 30 °C from 480 to 510 nm (slit width, 2 nm), using an emission wavelength of 545 nm (slit width, 20 nm) by a Synergy H1 plate reader (BioTek). Spectra were measured for three biological replicates at each pH. A standard curve was determined for green fluorescent protein by measuring fluorescence of samples resuspended in the same buffer at pH 5.5, 6.0, 6.5, 7.0 or 7.5 with addition of 20 mM sodium benzoate, a permeant acid equilibrating cytoplasmic pH with external pH.

In-line probing. Experiments were performed as described[39] with the following modifications: the *mgtC* leader RNA was synthesized *in vitro* with a T7 RiboMAX Large Scale RNA production system (Promega) from the DNA template amplified from wild-type 14028s and primers (8675/11562) for the *mgtC* leader 1–195. DNA templates with G_{95}→C, A_{44-46}→T or A_{56-57}→G substitutions were prepared by a two-step PCR reaction. For the first PCR reaction, we used two primer pairs 8675/8175 and 8174/11562 for G_{95}→C substitution, 8675/11727 and 11726/11562 for A_{44-46}→T substitution, 8675/11723 and 11722/11562 for A_{56-57}→G substitution and 14028s genomic DNA as a template. For the second PCR reaction, we mixed the two PCR products from the first PCR reaction as templates and amplified the DNA fragment with expected substitution using primers 8675 and 11562. To probe the structures at different Mg^{2+} concentrations, 1 pmol of 5′-end-labelled *mgtC* leader RNA was incubated in buffer (100 mM KCl, 50 mM Tris (pH 8.3)) with 1, 5 or 20 mM Mg^{2+} for 40 h at room temperature. Reactions were quenched with urea gel loading buffer II (Ambion) and analysed on a 10% denaturing polyacrylamide gel.

Examining the effect of pH on RNA structure. To determine pH effect on the structure of the *mgtC* leader, we used lead (II) acetate as it cleaves single-stranded RNAs non-specifically. One picomole of 5′-end-labelled wild-type 1–195 RNA or mutant RNA with A_{44-46}→T substitution was incubated in 10 µl of either pH 8.3 (100 mM KCl, 10 mM Tris (pH 8.3), 10 mM $MgCl_2$) or pH 7.0 (100 mM KCl, 10 mM Tris (pH 7.0), 10 mM $MgCl_2$) buffer for 10 min at 37 °C to allow RNA folding. Then, 1 µl of fresh solution of lead (II) acetate (50 mM) was added and incubated for 2 min at 37 °C. Reaction was stopped by adding 10 µl of gel loading buffer II (Ambion) on ice and analysed on a 10% denaturing polyacrylamide gel.

31. Fields, P. I., Swanson, R. V., Haidaris, C. G. & Heffron, F. Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *Proc. Natl Acad. Sci. USA* **83,** 5189–5193 (1986).
32. Davis, R. W., Bolstein, D. & Roth, J. R. *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory Press, 1980).
33. Snavely, M. D., Miller, C. G. & Maguire, M. E. The *mgtB* Mg^{2+} transport locus of *Salmonella typhimurium* encodes a P-type ATPase. *J. Biol. Chem.* **266,** 815–823 (1991).
34. Rust, M. J., Golden, S. S. & O'Shea, E. K. Light-driven changes in energy metabolism directly entrain the cyanobacterial circadian oscillator. *Science* **331,** 220–223 (2011).
35. Jensen, K. F., Houlberg, U. & Nygaard, P. Thin-layer chromatographic methods to isolate ^{32}P-labeled 5-phosphoribosyl-alpha-1-pyrophosphate (PRPP): determination of cellular PRPP pools and assay of PRPP synthetase activity. *Anal. Biochem.* **98,** 254–263 (1979).
36. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97,** 6640–6645 (2000).
37. Miller, J. H. *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, 1972).
38. Wilks, J. C. & Slonczewski, J. L. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J. Bacteriol.* **189,** 5601–5607 (2007).
39. Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. *Methods Mol. Biol.* **419,** 53–67 (2008).

doi:10.1038/nature11079

Atomic model of the type III secretion system needle

Antoine Loquet[1]*, Nikolaos G. Sgourakis[2]*, Rashmi Gupta[3]*, Karin Giller[1]*, Dietmar Riedel[4], Christian Goosmann[5], Christian Griesinger[1], Michael Kolbe[3], David Baker[2], Stefan Becker[1] & Adam Lange[1]

Pathogenic bacteria using a type III secretion system (T3SS)[1,2] to manipulate host cells cause many different infections including *Shigella* dysentery, typhoid fever, enterohaemorrhagic colitis and bubonic plague. An essential part of the T3SS is a hollow needle-like protein filament through which effector proteins are injected into eukaryotic host cells[3–6]. Currently, the three-dimensional structure of the needle is unknown because it is not amenable to X-ray crystallography and solution NMR, as a result of its inherent non-crystallinity and insolubility. Cryo-electron microscopy combined with crystal or solution NMR subunit structures has recently provided a powerful hybrid approach for studying supramolecular assemblies[7–12], resulting in low-resolution and medium-resolution models[13–17]. However, such approaches cannot deliver atomic details, especially of the crucial subunit–subunit interfaces, because of the limited cryo-electron microscopic resolution obtained in these studies. Here we report an alternative approach combining recombinant wild-type needle production, solid-state NMR, electron microscopy and Rosetta modelling to reveal the supramolecular interfaces and ultimately the complete atomic structure of the *Salmonella typhimurium* T3SS needle. We show that the 80-residue subunits form a right-handed helical assembly with roughly 11 subunits per two turns, similar to that of the flagellar filament of *S. typhimurium*. In contrast to established models of the needle in which the amino terminus of the protein subunit was assumed to be α-helical and positioned inside the needle, our model reveals an extended amino-terminal domain that is positioned on the surface of the needle, while the highly conserved carboxy terminus points towards the lumen.

Wild-type *S. typhimurium* T3SS needles were obtained by *in vitro* polymerization of recombinantly produced full-length PrgI protomers. The diameter of the resulting needle-like filaments (Supplementary Fig. 1b) agrees with that of T3SS needles from natural sources[3]. The addition of wild-type PrgI protomers also elongated needles of isolated *S. typhimurium* T3SSs up to several micrometres in length (Supplementary Fig. 2), showing the functionality of the recombinant protein (as previously reported for PrgI* (ref. 18)). Recently we demonstrated that excellent-quality solid-state NMR (ssNMR) spectra of PrgI needles can be obtained by ^{13}C spin dilution[19], which allowed us to achieve the complete ssNMR resonance assignment. As determined from conformation-dependent chemical shifts, PrgI adopts a rigid conformation (except for Met 1 and Ala 2) comprising four distinct structural elements (Fig. 1a): an N-terminal extended domain (Thr 3–Tyr 8), an α-helix (α1) (Leu 9–Ala 35), a loop (Ala 36–Pro 41) and a C-terminal α-helix (α2) (Ala 42–Arg 80). Very narrow ^{13}C linewidths (~0.15–0.3 p.p.m.) and the presence of a unique set of NMR signals exclude the possibility of conformational heterogeneity

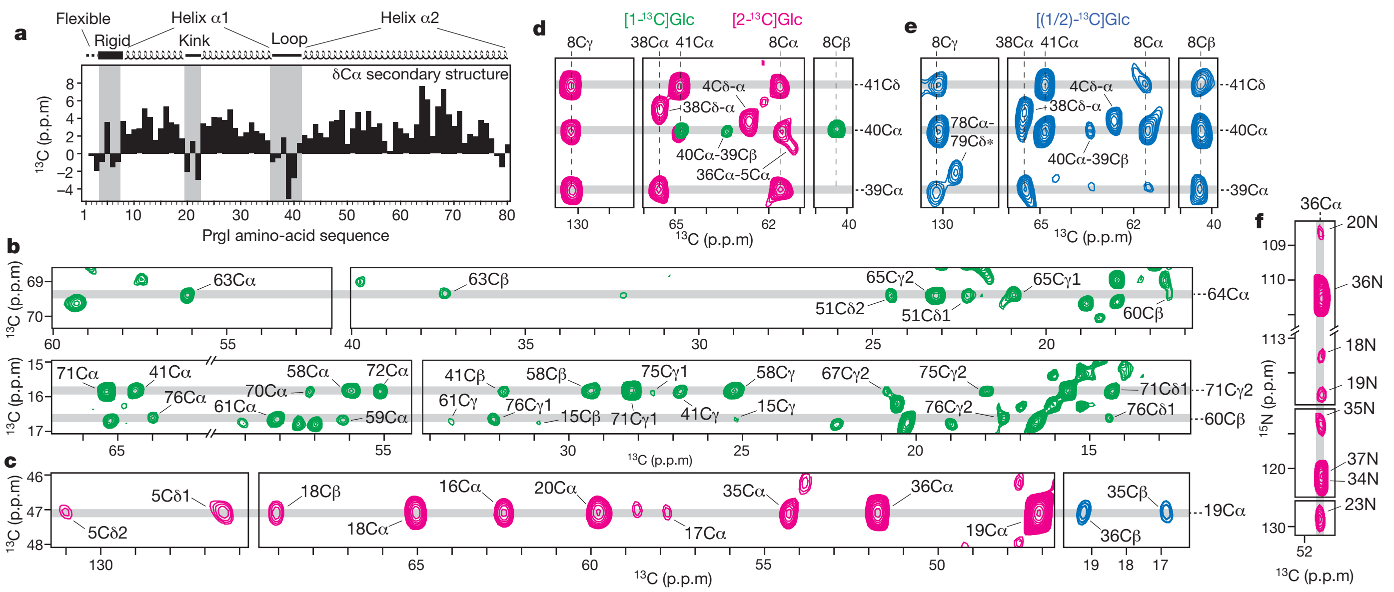


Figure 1 | Structural data from ssNMR. **a**, Secondary structure analysis. Positive and negative secondary chemical shifts are indicative of α-helical and β-sheet structure, respectively. **b**–**f**, Collection of ssNMR distance restraints. **b**, ^{13}C–^{13}C PDSD spectra of [1-^{13}C]Glc-labelled T3SS needles. **c**, ^{13}C–^{13}C spectra of [2-^{13}C]Glc-labelled (in magenta) and [(1/2)-^{13}C]Glc-labelled (in blue) T3SS needles. Intersubunit restraints can be obtained from the [(1/2)-^{13}C]Glc-labelled spectrum, for example, Gly 19Cα–Ala 35Cβ and Gly 19Cα–Ala 36Cβ. **d**, ^{13}C–^{13}C spectra of [1-^{13}C]Glc-labelled (in green), [2-^{13}C]Glc-labelled (in magenta) needles. **e**, ^{13}C–^{13}C spectra of [(1/2)-^{13}C]Glc-labelled (in blue) T3SS needles. Intersubunit restraints are encoded in the [(1/2)-^{13}C]Glc-labelled spectrum: Pro 41Cδ–Tyr 8Cβ and Ser 39Cα–Tyr 8Cβ. **f**, ^{15}N–^{13}C PAIN-CP spectrum of [2-^{13}C]Glc-labelled T3SS needles.

[1]Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [2]Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. [3]Department of Cellular Microbiology, Max Planck Institute for Infection Biology, 10117 Berlin, Germany. [4]Laboratory for Electron Microscopy, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. [5]Core Facility Microscopy, Max Planck Institute for Infection Biology, 10117 Berlin, Germany.
*These authors contributed equally to this work.

in the wild-type needle state, whereas in X-ray structures of soluble, C-terminally truncated protein[17] or the double-mutant PrgI* (V65A/V67A)[18] the first ~18 N-terminal residues are disordered (Supplementary Fig. 1c). In particular, ssNMR reveals a highly ordered N-terminal segment (Thr 3–Tyr 8) with unusual chemical shifts that do not correspond to regular α-helical or β-sheet conformation. A kink (Val 20–Asn 22) interrupts the helix α1 that is not observed in the monomeric X-ray crystal structures[18]. In PrgI* needles, β-strand conformation was observed at the C terminus[18] (Supplementary Fig. 1c), in contrast with the wild-type needle preparation. However, the observed structural differences could represent different functional states of the filament, as suggested previously[6].

The sparse [1-^{13}C]Glc or [2-^{13}C]Glc labelling significantly improves the quality of ^{13}C–^{13}C ssNMR spectra[19], and numerous cross-peaks can readily be assigned to long-range distance restraints (see Supplementary Figs 3–7 and Methods). The [1-^{13}C]Glc-labelled[20] sample (Fig. 1b) provides many methyl-based long-range restraints. Similarly, the [2-^{13}C]Glc-labelled[21] sample (Fig. 1c) shows excellent spectral resolution in the aromatic region, permitting the collection of unambiguous restraints. In total, 521 restraints were assigned (including 247 long-range ones; see Supplementary Table 1 and Methods), including long-range ^{15}N–^{13}C restraints (detected in a PAIN-CP (proton-assisted insensitive nuclei cross-polarization) spectrum[22]; Fig. 1f). Roughly twice as many unambiguous long-range restraints were found per residue than in the current benchmark in ssNMR structural studies, the HET-s prion domain[23].

To distinguish intrasubunit from intersubunit distance restraints, we compared spectra of [1-^{13}C]Glc-labelled, [2-^{13}C]Glc-labelled and mixed [(1/2)-^{13}C]Glc-labelled PrgI needle samples[24] (Fig. 1d, e) and identified 162 intersubunit and 359 intrasubunit interactions that together define the overall organization of the T3SS needle assembly (Fig. 2).

Intrasubunit restraints (Fig. 2, red lines) reveal a helical hairpin motif in PrgI. The kink segment (Val 20–Asn 22) disrupts the α1–α2 interactions, with no intrasubunit long-range restraints detected for residues Asp 17–Asn 22 and their counterpart Arg 58–Asn 59. The N-terminal part of α1 is again interacting with the C-terminal part of α2. Multiple restraints (Fig. 2, bold lines) were detected between Leu 9 and Lys 69, the most N-terminal and C-terminal residues that take part in the helical hairpin motif. No intrasubunit long-range restraints were detected for the N-terminal (Thr 3–Tyr 8) and C-terminal (Asp 70–Arg 80) segments, but the high intensity of their NMR signals[19] clearly demonstrates their rigidity and well-defined structure, indicating that these residues are stabilized through intersubunit contacts as described below.

The intersubunit distance restraints (Fig. 2, blue lines) reveal a lateral interface, mostly formed by helix–helix packing, and an axial interface. The lateral interface connects neighbouring PrgI subunits by means of α1–α1 and α2–α2 helix–helix packing in an arrangement in which the subunits are staggered along the filament axis with an axial translation of ~24 Å (indicated for example by restraints involving Ala 60 of one molecule and Ile 76 of an adjacent one). The last five C-terminal residues that are known to be essential for the needle formation and infectivity, in particular Ile 76 and Phe 79, are involved in multiple intersubunit restraints. The central part of α1 helices (Asp 17–Leu 23, including the kink) interacts with adjacent subunits through the extended N terminus, in particular the Trp 5 aromatic ring, and also through the end of helix α1 (Ala 35–Lys 37) of the adjacent subunit. The ssNMR restraints also reveal strong intersubunit interactions between the extended N-terminal domain (Thr 3–Tyr 8) and the loop region (Ala 36–Pro 41), defining the axial packing between subunits on top of each other (Fig. 2).

According to previous scanning transmission electron microscopy (STEM) measurements[25], the axial rise per subunit in PrgI needles is ~4.2 Å. The intermolecular distance restraints show that laterally adjacent subunits are axially translated from one turn to the next by

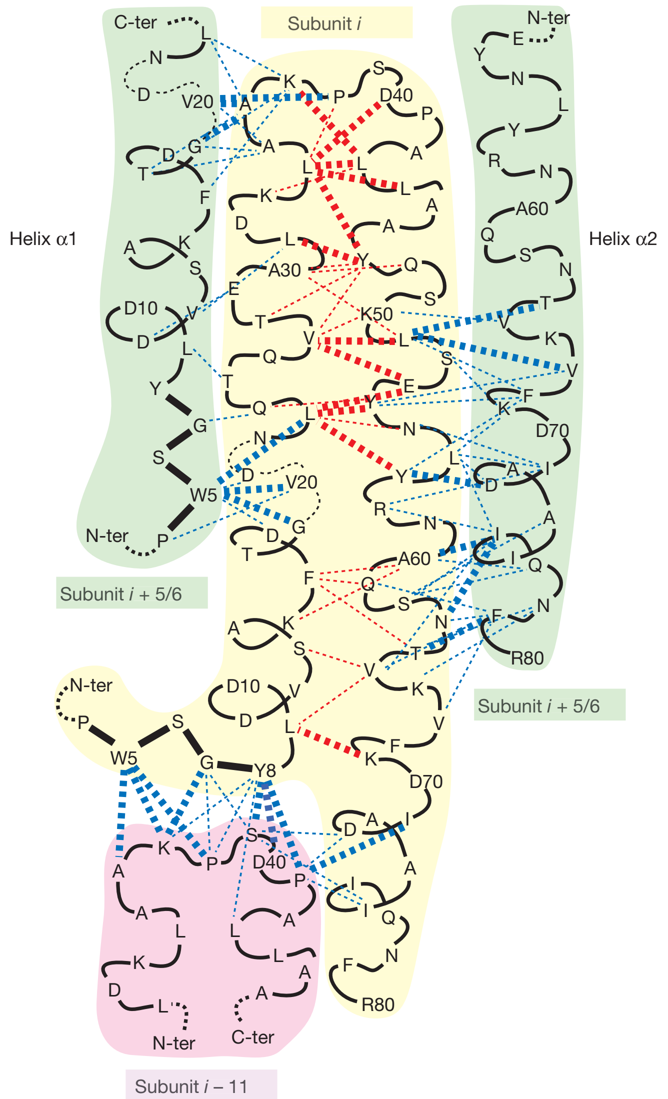


Figure 2 | Architecture of the T3SS needle assembly as determined by ssNMR. The dashed lines represent the intrasubunit (in red) and intersubunit (in blue) ssNMR distance restraints. Bold lines indicate the detection of more than three distance restraints. Lateral interactions between subunits *i* and *i* + 5 or *i* + 6 are not distinguishable at this stage of the analysis, but only after first rounds of Rosetta calculations (see Methods). N-ter, N terminus; C-ter, C terminus.

~24 Å, which divided by ~4.2 Å yields ~5.7 subunits per turn. These values are consistent with a helical arrangement in which roughly 11 subunits are arranged in two turns. To generate three-dimensional models of the filament, rather than docking preformed monomeric structures which would not capture possible conformational changes on assembly, we extended the Rosetta symmetric fold-and-dock protocol[26] to helical symmetry. Starting from a helically arrayed set of extended polypeptide chains, we optimized the total energy of the system including the ssNMR restraints and conformation-dependent chemical shifts by simultaneously sampling both the internal degrees of freedom of the monomers and the five rigid body degrees of freedom (the radius of the needle was restrained using the previously published cryo-electron microscopy map[25]). Models were generated for 9-start, 11-start, 13-start and 15-start helices (see Methods); the 11- start models had helical parameters more consistent with the STEM data[25] and had more favourable ssNMR restraint energies (Supplementary Table 3).

In an 11-start helical arrangement, interactions between subunits *i* and *i* ± 11 constitute the axial interface, and subunit *i* is laterally surrounded by subunits *i* ± 5 and *i* ± 6. We used first-round models

generated with only the unambiguous axial and intramolecular restraints to resolve the (5/6) ambiguities in the lateral restraints in two successive rounds of structure calculations (Supplementary Fig. 8). Whereas the inner α2 helices are parallel in the final models, the α1 helices present a cross-fenced pattern that emerges from the inclusion of two clusters of restraints, one involving Trp 5–Val 20 and the other Val 20–Ala 36 lateral interactions (Fig. 3; alternative assignments of these restraints resulted in significantly higher energy structures). The cross structure is further corroborated by the chemical shift data, which indicate a kink at Val 20 (Fig. 1a) as well as the lack of long-range intramolecular restraints for the same residue (Fig. 2). The combination of the ssNMR restraints with all-atom Rosetta refinement may yield near-atomic resolution; for example, the turn connecting the two helices has a backbone root mean squared deviation (r.m.s.d.) to the crystal structure of the monomer of less than 0.5 Å (Supplementary Fig. 8g).

The resulting T3SS needle model (Fig. 3 and Supplementary Fig. 9) consists of a ~80-Å outer diameter tube with a ~25-Å axial lumen

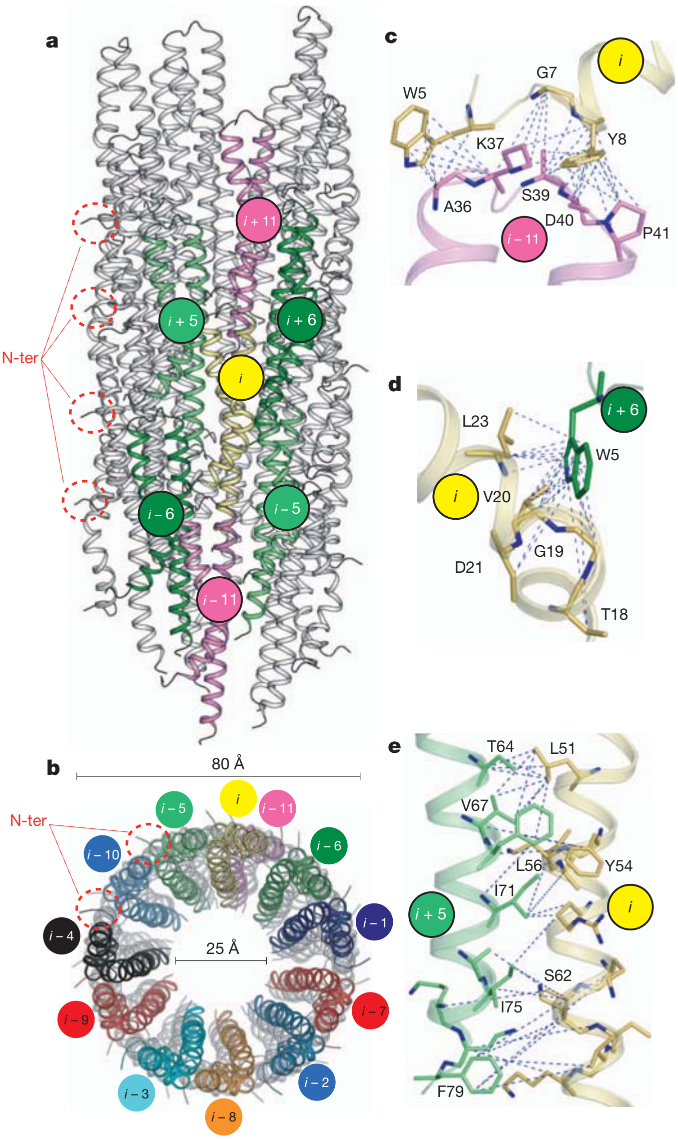


Figure 3 | Complete atomic model of the T3SS needle. **a**, **b**, Ribbon representation showing different subunits: side perspective (**a**); top view (**b**). The N-terminal domains (N-ter) are highlighted by red dashed circles. **c**–**e**, Selections of the subunit–subunit interfaces. **c**, The axial interface between subunits i and $i-11$. **d**, The lateral interface between subunits i and $i+6$. **e**, The lateral interface between subunits i and $i+5$. Blue dashed lines represent ssNMR restraints.

(Fig. 3b). The N-terminal domain was proposed to be α-helical in previous studies of the homologous *Shigella flexneri* needle[17,27] and to form the inner shell of the filament. Our atomic model of the T3SS needle clearly diverges from this prediction: the N-terminal domain is not α-helical and is located on the exterior surface of the needle, not on the inner side.

To corroborate this finding, we performed immunoelectron microscopy on both *S. typhimurium* cells and needles polymerized *in vitro* (Supplementary Figs 10 and 11). *S. typhimurium prgI-knockout* cells complemented with *prgI* fused to an upstream Strep-tag II were labelled with monoclonal anti-Strep antibodies (Supplementary Figs 11 and 13). The Strep-PrgI is fully functional in HeLa cell invasion assays of complemented *S. typhimurium prgI-knockout* cells (Supplementary Fig. 14). Similarly, *in vitro* needles of an N-terminal His-tag fusion with wild-type PrgI bind to an anti-His-tag monoclonal antibody that decorates the exterior surface of such needles (Supplementary Figs 10 and 12b). These results confirm that the PrgI N terminus is surface-exposed, an arrangement similar to that of the flagellar filament[14], which shares a genetic origin with the T3SS needle.

The rigid extended conformation of the N-terminal residues reported here clearly differs from previous models. The two aromatic residues Trp 5 and Tyr 8 form the basis of the axial N-terminal-domain–loop interface (Fig. 3c). Indeed, *in vivo* invasion assays show that the single point mutant Y8A leads to decreased invasiveness of *S. typhimurium* (Supplementary Fig. 14), and the double mutation W5A/Y8A completely abrogates host cell invasion (Supplementary Fig. 14). Trp 5 also seems to have a crucial function in the lateral i to $i+6$ interface (Fig. 3d; and W5A/V20A double mutant in Supplementary Fig. 14). Furthermore, the atomic model reveals that the lateral i to $i \pm 5$,6 interface strongly involves α2–α2 helix–helix contacts (Fig. 3e) and that the last C-terminal residues Ile 75, Ile 76 and Phe 79 are essential. This is confirmed by the non-invasiveness of the I76A/F79A PrgI double mutant (Supplementary Fig. 14). In addition, we compared *S. typhimurium* wild-type and *prgI-knockout* cells either complemented with wild-type or mutated *prgI* by using effector secretion assays and transmission electron microscopy (TEM) of a negatively stained specimen. The results obtained (Supplementary Fig. 16) agree well with the invasion assays shown in Supplementary Fig. 14. SDS–PAGE and western blot analysis showed that the double mutations (W5A/Y8A and I76A/F79A), but not the single mutations (W5A and Y8A), abolish effector secretion (Supplementary Fig. 16a, b). However, the W5A/V20A double mutant can secrete effector proteins, which is also in agreement with the results previously obtained by bacterial invasion assays. TEM analysis of these strains confirmed needles only in wild-type and *prgI-knockout* cells complemented with wild-type *prgI* (Supplementary Fig. 16c, d) but not in the mutant strains, which might be an indication of decreased needle stability.

Residues conserved between *S. typhimurium*, *Shigella flexneri*, *Yersinia pestis* and *Escherichia coli* T3SS needle proteins are mostly facing towards the lumen (Supplementary Figs 15 and 17). Weakly or non-conserved residues are exposed to the needle surface and are thus accessible from outside the pathogen, which could reflect a bacterial strategy to evade a host-cell immune response. The inner surface of the channel consists mainly of polar residues (see Supplementary Fig. 18 for the electrostatic surface potential of the T3SS channel). In a similar manner to the situation in the flagellar filament, the most C-terminal residue Arg 80 (in flagellin this is Arg 494 (ref. 14)) places a positive charge into the channel.

Our approach opens an avenue for the investigation of the subunit interfaces of other homo-oligomeric assemblies at atomic resolution in their native-like state[28–30]. For larger or more complex systems, such as the flagellar filament[14], our combination of atomic-level ssNMR structural restraints and cryo-electron microscopy data could bridge the gap between X-ray or NMR structures of monodisperse subunits and cryo-electron microscopy data of the filamentous state.

METHODS SUMMARY

Sample preparation. ^{15}N-labelled and ^{13}C-labelled PrgI protomers were expressed in *E. coli*. Polymerization of purified PrgI was allowed to take place at room temperature (22 °C) for 1 week.

Solid-state NMR. Experiments were conducted on 14.1-T and 20.0-T wide-bore Bruker spectrometers at magic angle spinning (MAS) frequencies of ~11 kHz. ^{13}C–^{13}C proton-driven spin-diffusion (PDSD) spectra were recorded with mixing times as indicated in the main text.

Structure calculation. T3SS needle structures were calculated with a modified Rosetta fold-and-dock protocol, using a 29-subunit system. The calculations converge within 2.1 Å backbone r.m.s.d.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 December 2011; accepted 23 March 2012.

Published online 20 May 2012.

1. Galan, J. E. & Wolf-Watz, H. Protein delivery into eukaryotic cells by type III secretion machines. *Nature* **444,** 567–573 (2006).
2. Cornelis, G. R. The type III secretion injectisome. *Nature Rev. Microbiol.* **4,** 811–825 (2006).
3. Kubori, T. *et al.* Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280,** 602–605 (1998).
4. Kimbrough, T. G. & Miller, S. I. Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation. *Proc. Natl Acad. Sci. USA* **97,** 11008–11013 (2000).
5. Tamano, K. *et al.* Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. *EMBO J.* **19,** 3876–3887 (2000).
6. Blocker, A. J. *et al.* What's the point of the type III secretion system needle? *Proc. Natl Acad. Sci. USA* **105,** 6507–6513 (2008).
7. Nogales, E. & Grigorieff, N. Molecular machines: putting the pieces together. *J. Cell Biol.* **152,** F1–F10 (2001).
8. Volkmann, N. & Hanein, D. Docking of atomic models into reconstructions from electron microscopy. *Methods Enzymol.* **374,** 204–225 (2003).
9. Rossmann, M. G., Morais, M. C., Leiman, P. G. & Zhang, W. Combining X-ray crystallography and electron microscopy. *Structure* **13,** 355–362 (2005).
10. Spreter, T. *et al.* A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nature Struct. Mol. Biol.* **16,** 468–476 (2009).
11. Baker, M. L., Zhang, J., Ludtke, S. J. & Chiu, W. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nature Protocols* **5,** 1697–1708 (2010).
12. Schraidt, O. & Marlovits, T. C. Three-dimensional model of *Salmonella*'s needle complex at subnanometer resolution. *Science* **331,** 1192–1195 (2011).
13. Fujii, T., Iwane, A. H., Yanagida, T. & Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* **467,** 724–728 (2010).
14. Yonekura, K., Maki-Yonekura, S. & Namba, K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424,** 643–650 (2003).
15. Wang, H. W. & Nogales, E. Nucleotide-dependent bending flexibility of tubulin regulates microtubule assembly. *Nature* **435,** 911–915 (2005).
16. Craig, L. *et al.* Type IV pilus structure by cryo-electron microscopy and crystallography: implications for pilus assembly and functions. *Mol. Cell* **23,** 651–662 (2006).
17. Deane, J. E. *et al.* Molecular model of a type III secretion system needle: implications for host-cell sensing. *Proc. Natl Acad. Sci. USA* **103,** 12529–12533 (2006).
18. Poyraz, O. *et al.* Protein refolding is required for assembly of the type three secretion needle. *Nature Struct. Mol. Biol.* **17,** 788–792 (2010).
19. Loquet, A., Lv, G., Giller, K., Becker, S. & Lange, A. ^{13}C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. *J. Am. Chem. Soc.* **133,** 4722–4725 (2011).
20. Hong, M. Determination of multiple ϕ-torsion angles in proteins by selective and extensive ^{13}C labeling and two-dimensional solid-state NMR. *J. Magn. Reson.* **139,** 389–401 (1999).
21. Lundstrom, P. *et al.* Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]- or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins. *J. Biomol. NMR* **38,** 199–212 (2007).
22. Lewandowski, J. R., De Paepe, G. & Griffin, R. G. Proton assisted insensitive nuclei cross polarization. *J. Am. Chem. Soc.* **129,** 728–729 (2007).
23. Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218–289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319,** 1523–1526 (2008).
24. Loquet, A., Giller, K., Becker, S. & Lange, A. Supramolecular interactions probed by ^{13}C–^{13}C solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **132,** 15164–15166 (2010).
25. Galkin, V. E., Schmied, W. H., Schraidt, O., Marlovits, T. C. & Egelman, E. H. The structure of the *Salmonella typhimurium* type III secretion system needle shows divergence from the flagellar system. *J. Mol. Biol.* **396,** 1392–1397 (2010).
26. Das, R. *et al.* Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl Acad. Sci. USA* **106,** 18978–18983 (2009).
27. Kenjale, R. *et al.* The needle component of the type III secreton of *Shigella* regulates the activity of the secretion apparatus. *J. Biol. Chem.* **280,** 42929–42937 (2005).
28. Goldbourt, A., Gross, B. J., Day, L. A. & McDermott, A. E. Filamentous phage studied by magic-angle spinning NMR: resonance assignment and secondary structure of the coat protein in Pf1. *J. Am. Chem. Soc.* **129,** 2338–2344 (2007).
29. Han, Y. *et al.* Solid-state NMR studies of HIV-1 capsid protein assemblies. *J. Am. Chem. Soc.* **132,** 1976–1987 (2010).
30. Jehle, S. *et al.* Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nature Struct. Mol. Biol.* **17,** 1037–1042 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. C. Marlovits and E. H. Egelman for providing the *S. typhimurium* T3SS needle cryo-electron microscopy density map; F. DiMaio and J.-P. Demers for discussions; and G. Wolf, B. Angerstein and G. Heim for technical help. This work was supported by the Max Planck Society (to C. Griesinger), the Deutsche Forschungsgemeinschaft (Emmy Noether Fellowship to A. Lange), the Fondation Bettencourt Schueller (to A. Loquet), the National Institutes of Health (1 R01 GM092802-01 to D.B.), EMBO (postdoctoral fellowship to A. Loquet), and the European Union Seventh Framework Program under Grant Agreement 261863 (Bio-NMR).

Author Contributions A. Loquet performed ssNMR experiments. A. Loquet and A. Lange analysed ssNMR data. N.S. and D.B. performed structure calculations. K.G. and S.B. expressed, purified and polymerized *in vitro* T3SS needles. R.G. and M.K. performed the *in vivo* studies. C. Griesinger analysed NMR data. D.R. and C. Goosmann performed electron microscopy studies. A. Loquet and A. Lange wrote the paper; all authors discussed the results and commented on the manuscript.

Author Information The *Salmonella typhimurium* T3SS PrgI needle structure is deposited in the Protein Data Bank under the accession code 2LPZ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A. Lange (adla@nmr.mpibpc.mpg.de), S.B. (sabe@nmr.mpibpc.mpg.de) or M.K. (kolbe@mpiib-berlin.mpg.de).

METHODS

Expression, purification and polymerization of ^{15}N-labelled and ^{13}C-labelled wild-type PrgI protein. The synthetic full-length PrgI encoding DNA was purchased from GeneArt and cloned into a modified pET16b vector (Novagen). The resulting construct codes for a fusion protein with an N-terminal His_7 tag followed by a tobacco etch virus (TEV) protease cleavage recognition sequence that connects it to the PrgI protein sequence. Thus, after protease cleavage the released PrgI protein contains the non-native N-terminal residues glycine and histidine. The construct was transformed into *E. coli* strain BL21(DE3). Expression of labelled PrgI protein was performed in minimal medium with $^{15}NH_4Cl$ as nitrogen source and D-[U-$^{13}C_6$]glucose, D-[1-^{13}C]glucose or D-[2-^{13}C]glucose as carbon source. An overnight minimal medium culture was inoculated 1:25 (v/v) into 1 l of minimal medium. The culture was shaken at 37 °C until a $D_{600\,nm}$ of 0.7. Protein expression was induced by adding 0.7 mM isopropyl β-D-thiogalactoside; 5 h after induction the cells were harvested and stored at −80 °C.

The cell pellet of a 1-litre expression culture was dissolved in 50 ml of lysis buffer (100 mM sodium phosphate, 10 mM Tris-HCl, pH 8.0, 8 M urea, 0.5 mM phenylmethylsulphonyl fluoride (PMSF)) and stirred for 1 h at room temperature. Subsequently, the suspension was centrifuged for 30 min at 33,000*g* and the supernatant was incubated on a rocking shaker for 1 h with 3 ml of Ni^{2+}-nitrilotriacetate (Qiagen) resin that had been pre-equilibrated with the lysis buffer. The resin suspension was filled into a 10-ml disposable plastic column (Pierce). The column was washed with 60 ml of lysis buffer and the protein was eluted with 8 × 2 ml of elution buffer 1 (100 mM sodium phosphate, 10 mM Tris-HCl, pH 5.9, 8 M urea, 0.5 mM PMSF) and 8 × 2 ml of elution buffer 2 (100 mM sodium phosphate, 10 mM Tris-HCl, pH 4.5, 8 M urea, 0.5 mM PMSF). The elution fractions were analysed on a 17.5% SDS–PAGE gel. Fractions containing the fusion protein were combined and concentrated with a Vivascience 5-kDa molecular mass cut-off concentrator (Sartorius Stedim Biotech) to a final volume of 4 ml. Subsequently, 2-ml aliquots of the sample were loaded onto a 8 mm × 250 mm semipreparative HPLC column (RP18 Eurospher 100; Knauer) that had been pre-equilibrated with 0.1% aqueous trifluoroacetic acid (TFA). The column was washed with ten column volumes of 0.1% TFA, and the protein was eluted at a flow rate of 3 ml min^{-1} with a 150-ml linear gradient of 0–100% acetonitrile/0.1% TFA. The elution peaks were analysed by electrospray ionization mass spectrometry (ESI-MS). Elution fractions containing the fusion protein without any further detectable impurities were combined and freeze-dried.

The freeze-dried protein was dissolved in ice-cold water by vortex-mixing and then incubated for 1 h on ice. Subsequently, using cold stock solutions, the following buffer was adjusted in this aqueous protein solution: 50 mM Tris-HCl, pH 8.0, 0.5 mM EDTA, 1 mM dithiothreitol, 0.5 mM PMSF. To this buffered protein solution TEV protease was added at a final ratio of 3 mg of TEV protease per 100 mg of fusion protein. The digestion was performed for 20 h at 4 °C. The released PrgI protein was further purified by reverse-phase HPLC as described above for the fusion protein. Elution fractions containing pure PrgI protein according to ESI-MS analysis were combined and freeze-dried.

The freeze-dried protein was dissolved at 0.085 mM concentration ($\varepsilon_{280} = 10{,}810\ M^{-1}\ cm^{-1}$) in ice-cold water by vortex-mixing and dialysed at 4 °C against 20 mM MES, pH 5.5, 0.02% sodium azide. Finally, the sample was concentrated with a Vivascience 5-kDa molecular mass cut-off concentrator to a final concentration of 0.1 mM.

The polymerization of the PrgI protein took place by incubation at room temperature for 1 week without shaking the sample. Thereafter the needles were centrifuged for 30 min at 52,000*g*. The protein pellet was washed extensively with 20 mM MES pH 5.5 before transfer to the ssNMR rotor.

Invasion assays. HeLa cell invasion assays were performed as described in ref. 18. In brief, cells were infected with *S. typhimurium* SL1344 and PrgI mutants at multiplicity of infection of 10:1 after inducing their expression with 0.2 μg ml^{-1} anhydrotetracycline for 1 h. After 20 min of infection, the cells were washed with PBS, and fresh medium supplemented with 100 μg ml^{-1} gentamicin was added. Cells were incubated for a further 2 h to allow intracellular survival and replication of bacteria. They were then lysed with 0.1% Triton X-100 and the lysates were plated on tryptic soy agar plates to count the number of viable bacteria after 16 h.

Secretion assays. Single colonies of *S. typhimurium* SL1344 and PrgI mutants were inoculated in Luria–Bertani medium with 0.04 μg ml^{-1} anhydrotetracycline and incubated with shaking at 37 °C for 16 h. The cultures were centrifuged at 4,354*g* for 10 min to remove the bacteria. The supernatants were filtered through 0.22-μm filters and precipitated with trichloroacetic acid. Pellets were washed once with acetone, dried, resuspended in SDS loading buffer and analysed by SDS–PAGE.

Preparation of cells for electron microscopy analysis. Cultures of *S. typhimurium* were grown as described above, and 1 ml aliquots were fixed with 2% paraformaldehyde for 1 h. The fixed cells were washed once with PBS and resuspended in 200 μl of PBS for electron microscopy analysis.

ssNMR spectroscopy and detection of distance restraints. ssNMR experiments were conducted on 14.1-T and 20.0-T (^{1}H resonance frequencies of 600 and 850 MHz, respectively) wide-bore spectrometers (Bruker Biospin) equipped with 4 mm triple-resonance (^{1}H, ^{13}C and ^{15}N) MAS probes. Samples were packed in 4-mm MAS rotors, using protein quantities of ∼10 mg. Spectra were recorded at MAS spinning rates of 10.75 or 11 kHz and were calibrated with 4,4-dimethyl-4-silapentane-1-sulphonic acid (DSS) as an internal reference. The temperature-dependent position of the water proton resonance was used to measure the temperature inside the MAS rotor[31]. All experiments were conducted at a sample temperature of 278 K. High-power ^{1}H–^{13}C decoupling (SPINAL-64)[32] with a radiofrequency amplitude of 83 kHz was applied during evolution and detection periods. An initial ramped cross-polarization was used to transfer magnetization from ^{1}H to ^{13}C with a contact time of 1.5 ms. ^{13}C–^{13}C polarization transfer was achieved by means of PDSD[33] with mixing times as indicated in the text and in the supplementary figures.

A ^{15}N–^{13}C PAIN-CP spectrum was recorded as described in ref. 19. Total experimental times were 4–9 days depending on the number of scans (maximum acquisition times t_1 of 12 ms for [1-^{13}C]Glc-labelled and [2-^{13}C]Glc-labelled samples, and 7 ms for the mixed-labelled sample were used). Distance restraints with a unique assignment (spectrally unambiguous) were collected in a first round of analysis (corresponding to ∼70% of the total number of restraints) and used to clarify the assignment of additional ambiguous restraints (with four or fewer assignment possibilities). Intermolecular distance restraints were first assigned in the [(1/2)-^{13}C]Glc-labelled sample by direct comparison with [1-^{13}C]Glc-labelled and [2-^{13}C]Glc-labelled samples, as described in ref. 24. On the basis of this first set of restraints, additional intermolecular distances were later identified in spectra of the [1-^{13}C]Glc-labelled and [2-^{13}C]Glc-labelled samples.

NMR spectra were analysed with CcpNmr[34].

Modelling. A modification of the fold-and-dock protocol[26], now part of the standard Rosetta 3.3 distribution, was used for all model calculations. All symmetric subunits within the needle helix are treated explicitly in a 29-subunit system. The protocol consists of backbone fragment insertion trials that are replicated among symmetric subunits and symmetrical rigid body trials of adjacent subunits using a coarse-grain representation of the system, followed by side-chain packing and all-atom refinement of the side chain, backbone and rigid body degrees of freedom, using a physically relevant energy function[35].

The chemical shift assignments of the backbone ^{15}N and ^{13}C atoms from ssNMR experiments were used to bias the selection of overlapping three-residue and nine-residue fragments of backbone conformations. Using these fragments and the intrasubunit and intersubunit restraints, we performed a series of model calculations, starting from a helix of extended chains. A pseudo-energy term was used to restrain the calculations according to the agreement with the experimental distances, while optimizing the Rosetta energy function. ssNMR cross-peaks were converted to internuclear C–C and C–N distance restraints with an upper limit of 5 Å and a penalty that grew exponentially with distance.

Intersubunit distance restraints were first classified into two categories: the 'axial' restraints corresponding to i to $i \pm 11$ (equivalently i to $i \pm n$ for an n-start arrangement in the more general case) intersubunit interactions and the 'lateral' restraints corresponding to i to $i \pm 5/6$ intersubunit interactions. The resolution of (5/6) ambiguities was made through the manual identification of self-consistent clusters of contacts between subunits i and $i + 5$ or between subunits i and $i + 6$.

In detail, this was done in three stages. In stage I, using the axial and intrasubunit restraints alone, we identify two clusters of helical arrangements (Supplementary Fig. 8b). The first cluster corresponds to a left-handed helix of 5.6 subunits per turn that showed lateral interactions between subunits i to $i + 5$ and i to $i + 6$, and the second cluster is a right-handed helix with 11 subunits per turn that shows lateral interactions between subunits i to $i - 1$ and i to $i + 1$. This model is in disagreement with the estimates of subunits per turn of ∼5.6 from STEM and ssNMR data (see the text). We therefore focus on the first cluster, using which we can easily assign two sets of lateral ssNMR restraints that report on the i to $i + 5$ and i to $i + 6$ interfaces between the inner (C-terminal) helices (Supplementary Fig. 8a). Using the axial, intrasubunit and the stage I assigned lateral restraints we perform a second set of calculations, one for each helical handedness as the same lateral interactions can be satisfied by a right-handed helix and by a left-handed helix (stage IIR and stage IIL). This is done by keeping the axial and intersubunit interactions fixed but interchanging the restraints assigned for the i to $i + 5$ and i to $i + 6$ interfaces (Supplementary Fig. 8d). Using this second round of models we can further assign two clusters of restraints that correspond to lateral interactions connecting the outer helices (Supplementary Fig. 8c). This corresponds to two clusters of interactions involving W5 from subunit $i + 11$ and A36 from subunit i with the V20 kink region of subunit $i + 6$ in the left-handed model, shown with red and yellow dashes, respectively, in Supplementary Fig. 8c. A final stage of calculations is then performed using all assigned lateral restraints for each helix

(right-handed and left-handed) and the resulting ensembles are reported as the final models. The calculations converge within 2.1 Å backbone r.m.s.d. for the right-handed and 2.3 Å r.m.s.d. for the left-handed model (Supplementary Fig. 8e, f). Both models satisfy the experimental restraints with a minimal number of violations, although the right-handed model shows more favourable restraint and interface energies. Furthermore the two ensembles show very similar helical parameters (number of subunits per turn, radius and axial rise) and also full-atom interface energies (Supplementary Table 2). We note that the central part of the needle protein (Leu 34–Leu 43) is highly conserved between our atomic model and the crystal structure of PrgI* (ref. 18) (Supplementary Fig. 8g).

The radius of the needle was restrained during each calculation by using a pseudo-energy function that measures the correlation of the whole model to the previously published cryo-electron microscopy map[25]. In all calculations, no assumptions were made about the helical parameters of the needle (that is, the degree of angle rotation and unit translation along the helical axis). Instead, these parameters emerge in the final structural ensemble according to the interpretation of the intermolecular interactions. Calculations are repeated for different interpretations of the intersubunit distance restraints according to 9-start, 13-start and 15-start helical arrangements in addition to the 11-start presented here. These complementary calculations result in converged structural ensembles with similar axial and lateral interfaces but involving interactions between different symmetry-related monomers and helical parameters (Supplementary Table 3). Although no unambiguous distinction can be made between these geometries in terms of agreement with the ssNMR distance restraints alone, the symmetry considerations presented in the main text suggest that the 11-start arrangement agrees better with the complementary STEM measurements. This was the basis for selecting the 11-start needle as the final model.

31. Bockmann, A. *et al.* Characterization of different water pools in solid-state NMR protein samples. *J. Biomol. NMR* **45,** 319–327 (2009).
32. Fung, B. M., Khitrin, A. K. & Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **142,** 97–101 (2000).
33. Szeverenyi, N. M., Sullivan, M. J. & Maciel, G. E. Observation of spin exchange by two-dimensional Fourier-transform C-13 cross polarization-magic-angle spinning. *J. Magn. Reson.* **47,** 462–475 (1982).
34. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696 (2005).
35. Bradley, P., Misura, K. M. & Baker, D. Toward high-resolution *de novo* structure prediction for small proteins. *Science* **309,** 1868–1871 (2005).

CORRECTIONS & AMENDMENTS

RETRACTION

doi:10.1038/nature11164

Retraction: DNA demethylation in hormone-induced transcriptional derepression

Mi-Sun Kim, Takeshi Kondo, Ichiro Takada, Min-Young Youn, Yoko Yamamoto, Sayuri Takahashi, Takahiro Matsumoto, Sally Fujiyama, Yuko Shirode, Ikuko Yamaoka, Hirochika Kitagawa, Ken-Ichi Takeyama, Hiroshi Shibuya, Fumiaki Ohtake & Shigeaki Kato

Nature **461,** 1007–1012 (2009); doi:10.1038/nature08456.

In this Letter, we claimed that hormone-regulated transcriptional control involves DNA demethylation mediated by MBD4. Subsequently, we corrected some figure panels that appeared to have been erroneously prepared (Corrigendum, http://www.nature.com/nature/journal/v480/n7375/full/nature10604.html). However, we later found that other figure panels contained data duplications. After further review, we now conclude that the results presented in the original figures had been inappropriately manipulated, and given these more serious concerns we no longer have confidence in the original figures. We therefore wish to retract this Letter and sincerely apologise for any adverse consequences that may have resulted from the paper's publication.

FUTURES SCIENCE FICTION

SQUEALER

Mouthpiece for a generation.

BY ROBERT NATHAN CORRELL

JACEY

"Hey George!"

A blast of cool, processed air entered the room as my new boss strolled in.

I stopped typing. "The name's Ben."

"Not today." He threw another requisition on top of the towering stack of papers already on the desk. "New request just came in over the squealer. It's a priority — something about farm animals."

I glanced over the sheet. "Sure. Just give me whatever sample you've got of this guy's writing and I'll get on it as soon as I finish this."

"We don't have anything. Just a list of titles and a name. This guy's stuff must have got lost at the beginning of the Big Crash. There are no samples. Go wild. And do it now. That," he indicated my current project, "can wait. Two days."

"Two? I usually have at least four to pound one of these out."

"Not this time. I told you — it's urgent."

"Looks like he did a book on livestock and what else, some kind of single-year history? What's the rush?"

The boss sighed. "Some idiot let an old beta version of the Omnipedia out on the web and there are links to this guy all over it. Download requests are hitting us from e-readers all over the country. You're it, George."

"Hey, Ben — the guys and I are going out for a drink to catch the daily newsview. Want to come?"

I stopped typing and sighed. "Not today, Muriel. Priority request. Boss says he needs it yesterday."

"Seriously? Look at you — sweating it out in this little closet on that ancient P. O. S. all damn day … why do you put up with it?"

"Boss says it's verisimilitude. You guys all get to use those telepathic scribes, but that doesn't sound the same as what you get out of a typewriter. Something about cadence." I shook my head. "Besides, those things record everything that goes through your head. Doesn't that bother you?"

She laughed. "Thinking seditious thoughts? Or just trying to make out your grocery list for the week? Whatever — this is a dead-end job. You should ask for a promotion to tech writing or something. You could get an office next to mine upstairs. At least we don't have to dig through these piles of paper." She prodded the swaying stacks of copy with disgust.

I shook my head. "You all go on. I'll catch up with you later. Maybe Wednesday."

"You really love this crap, don't you? The deadlines, the typewriter, the stories … If you weren't just churning out those prop pieces about duty to state and the virtues of hard work, I'd think you were the only real writer left in the country."

"Don't say that, Muriel." I looked out the door with an exaggerated motion and smiled. "Someone could hear."

That night I walked back to the apartment, past shuttered stores and boarded-up old tenements. It was still long before curfew and some people were about. A few were entering a dimly lit dive to catch the three hours of newsview that were mandated each night, while others tapped and scratched at their electronic tablets. I wondered if they were reading anything of mine. Maybe some Burroughs or Asimov — those had been fun to do. I rubbed my head. This afternoon had not gone so well and I wished I had gone with Muriel and the others after all. The words just wouldn't come. Not writer's block, but something about my assignment's title that tickled the back of my mind and wouldn't let go.

After an hour or so of digging through the hidden compartment under my bed, I found the answer. It was in the stack of old volumes I picked out of my grandmother's storage unit the year before. A decaying paperback with a pig and a horse on the cover, barely visible beneath mould and water stains. I held it, running a finger across the pages and watching them flake away. What a trick — a bit of sleight of hand, electronic legerdemain. The boss once told me: "Just too much was lost in the Crash. So much of our history. And the idea that it's just gone would be too much for most people to believe." I looked at the book in my hand and laughed. I was replacing entire texts with real bits of fiction, and there was no one to know the difference but a dying generation and their failing memories. And me. I looked around my nice place. A real bed. Kitchenette. My own bathroom. Even Muriel didn't have one of those. On the newsview, a government drone dropped bombs on some Nomad shacks, while the occupants were marched away at rifle-point. Re-settlement, re-education. But I never saw a man who could write a good propaganda piece in a work camp. And I meant to be writing for a long, long time. Turning, I pushed the book down the incinerator chute and began heading back to the office, off to write a young adult novel about the exciting adventures of a patriotic farmer and his loyal farm creatures, all toiling for the greater good. This, for instant delivery to the tablet computers of would-be revolutionaries and faux counterculture icons all over the nation, each of them looking for the ideas that could change their world. ■

NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mtoodm

Robert Nathan Correll *is a postdoctoral fellow in cardiovascular biology who lives in Kentucky. He does not own an e-reader.*

CAREERS

TURNING POINT Molecular biologist tells how he's held on to his grant for 30 years **p.283**

CAREERS ADVICE FORUM Get expert advice on science careers issues **go.nature.com/lm1x4t**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

IKON IMAGES/CORBIS

PHARMACEUTICAL SECTOR

Delicate transition

With lay-offs rife in the drug industry, life scientists and chemists are seeking fresh career paths.

BY CHARLOTTE SCHUBERT

Locals were irate when the drug giant Pfizer closed its 70-hectare research and development (R&D) facility in Ann Arbor, Michigan, in 2007. T-shirts sporting the word 'Pfired' appeared on the streets; the governor called the lay-offs a "punch to the gut"; and the state of Michigan pledged US$1 million to help the 2,100 displaced workers find new jobs.

The pharmaceutical industry has faced major upheaval in recent years, with a disappointing drug pipeline, major revenue losses as patents expire on blockbuster drugs, and a spate of mergers and acquisitions. From 2006 through to the first quarter of 2012, some 263,000 positions have been eliminated from major pharmaceutical and large biotechnology companies, says Kenneth Getz, a senior research fellow at the Tufts Center for the Study of Drug Development in Boston, Massachusetts. R&D operations have accounted for 7–10% of the lay-offs since 2008, which have been only partially offset by new hiring, endangering what was once a stable source of jobs for life scientists and chemists.

But despite such convulsions, there are positive signs in the job market. In the Ann Arbor region, dozens of contract-research organizations (CROs), many founded by former Pfizer employees, offer outsourced services ranging from medicinal chemistry to toxicology testing. The abandoned Pfizer facility has been reborn: the University of Michigan bought it and now uses some of it as research facilities and rents out another part to Lycera, a biotechnology spin-off from the university that partners with pharmaceutical company Merck and employs some former Pfizer scientists.

"What we are seeing in front of our eyes is the slow-motion implosion of the big pharma companies as we know them, and the rebirth of the industry with different models and in different forms," says Bernard Munos, founder of the InnoThink Center for Research in Biomedical Innovation in Indianapolis, Indiana.

Researchers looking for work in this environment need to adapt their skills to an industry in flux, says Munos, and consider how to use their experience to secure a new type of job. They should also be aware that laid-off researchers may have to take jobs at lower salaries at CROs or biotechnology start-ups, or in other industries. In short, to weather the cuts — which show no signs of abating — pharma employees and new graduates "are going to have to hustle", says Munos.

R&D BREAKDOWN

In the face of declining pharmaceutical revenues, a variety of strategies has emerged to increase productivity and decrease costs. Companies such as GlaxoSmithKline, headquartered in London, have broken up research departments into smaller, more nimble units, and many firms are outsourcing R&D that would once have been done in-house. Meanwhile, research areas have been cut back. For example, Novartis, based in Basel, Switzerland, has reduced development of drugs that affect the central nervous system, considered a high-risk, expensive field (see *Nature* **480,** 161–162; 2011).

Firms, including Novartis, are shifting ▶

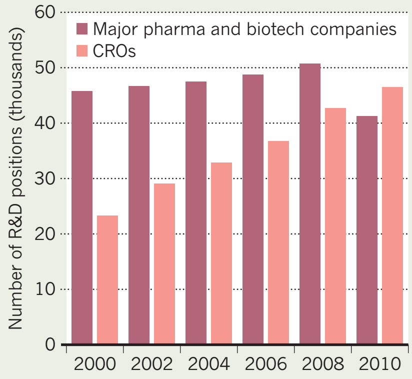
operations to areas such as Boston, where they can mine academia and biotech companies for early-stage discoveries, and China, an emerging market with a growing scientific workforce. Outsourcing may account for much of the net workforce reductions over the past several years, which occurred even as total investment in R&D by major pharmaceutical and biotechnology companies increased. An estimated 41,275 workers were employed in pharmaceutical- and biotechnology-industry R&D worldwide in 2010, down from 50,750 in 2008, according to the Tufts Center (see 'Outsourcing on the rise').

The closures affect all workers, from laboratory heads to technicians. But some jobs seem to be more vulnerable than others. When Pfizer, which is based in New York City, laid off employees in Ann Arbor, it offered jobs to hundreds of them at other locations. Most were scientists with transferable skills, such as computational biologists, or worked in hot areas such as oncology, says John LaMattina, who oversaw the lay-offs as head of global R&D at Pfizer and is now a senior partner at Puretech Ventures, a life-sciences venture-capital company in Boston. Specialists in waning fields are often most vulnerable, he says.

Bench scientists who work in the earliest stages of drug research may also be at high risk, as many pharmaceutical companies turn to academia and biotechnology companies for leads. "Major R&D organizations within big pharma have just been slashing without a lot of regard in drug discovery," says John Archer, founder of Catalyst Advisors in New York City, which recruits executives for pharmaceutical and biotechnology companies. People who work in clinical research and regulatory affairs seem to be better buffered from lay-offs, he says.

OUTSOURCING ON THE RISE

As pharma downsizes, contract-research organizations (CROs) have added workers.



TUFTS CSDD.

STAYING AHEAD

It is difficult to trace where the jobs are going, but a lot of people do manage to find work. In the United Kingdom, for instance, about 2,000 chemists at pharmaceutical companies were laid off last year, estimates Charlotte Ashley-Roberts, a careers adviser at the Royal Society of Chemistry, based in Cambridge, UK. More than 85% found jobs within three months, and 60% in chemistry. "You are starting to see a whole different cadre of opportunities," says John Arrowsmith, a life-sciences adviser at Thomson Reuters in London.

Many former pharmaceutical researchers are heading to CROs (see 'Prepare and contrast'), which have been growing steadily in the United States, Europe, India and China in recent years (see *Nature* **466,** 280–281; 2010). In 2010, 46,550 people were employed in R&D at CROs worldwide, up from 42,687 in 2008, estimates the Tufts Center.

People with skills beyond bench work are moving into consultancy, as experts in areas such as regulatory affairs, clinical-trial management and biostatistics (see 'Use your skills'). But there are no hard numbers on who is taking this route — or on how many consultants are effectively underemployed.

Yet other workers are retooling their skills for related industries that remain strong, including development of medical food (such as 'gut-healthy' yogurt), medical-device engineering and biomanufacturing, says Clifford Minz, founder of BioInsights, a career consultancy in Princeton, New Jersey. Patent specialists and medical writers are also in demand, he adds.

THE GREAT LEAP SIDEWAYS

Alex Flood is a former pharma researcher who has successfully made the transition to one growing niche sector: non-profit work. He "cut his teeth" at Wyeth and weathered that company's 2009 buy-out by Pfizer, but for many years he had aspired to a job in public health. Since 2010, he has been employed at PATH, a global-health non-profit organization in Seattle, Washington, where he works on vaccine stabilization — by, for example, devising ways to keep vaccines fresh over time. To find new work, "you have to be flexible", says Flood, who adapted his pharma training to his new job.

Peter Corr, co-founder of Celtic Therapeutics, a private-equity drug-development firm in New York City, hires senior and junior pharmaceutical professionals with a wide range of experience, from outsourcing to finance. He was head of science and technology at Pfizer until 2006, and says that it helps if candidates have an understanding of the whole drug-development pipeline. "Spend some time in your off hours in other parts of the company," he advises potential applicants. For example, bench scientists should expand their skill sets by learning about regulatory affairs or business development. Being open to relocation also helps, says Archer, given the geographic shifts in the industry.

The best candidates show passion for what they do and have taken on challenges, says Corr. "You see people who are moving to gain new experiences," he says. "These people are constantly stretching themselves." ■

Charlotte Schubert *is a freelance journalist based in Seattle, Washington.*

INTO CONTRACT RESEARCH

Prepare and contrast

Contract-research organizations (CROs) are increasingly taking on tasks previously done within the pharmaceutical industry, from clinical-trial management to medicinal chemistry. Many former employees of big pharmaceutical companies are moving to CROs — and liking it. "I thought the environment was going to be significantly different," says Jim Kremidas, who made the leap in 2008. But the change "was not as traumatic as I thought it was going to be".

Kremidas had spent more than 20 years in various jobs at Eli Lilly in Indianapolis, Indiana, culminating in a role as head of patient recruitment. But as the company downsized in preparation for patent expirations, he accepted a severance package. Knowing what lay ahead, Kremidas had time to prepare.

Before his retirement, Kremidas began developing his contact network. He also took on speaking engagements through the Drug Information Association, an international industry trade group based in Horsham, Pennsylvania. The exposure helped to build his reputation as a leader in patient recruitment, he says, and in 2008 he was offered a job doing just that at Quintiles, a CRO based in Durham, North Carolina. This February, Kremidas became head of market development for digital patient recruitment at the firm, using the Internet and online patient databases to gather volunteers.

Kremidas advises pharmaceutical employees who are worried about lay-offs to broaden their experiences to prepare themselves for a variety of future challenges. "I knew I had a skill set that was needed in industry, and Quintiles seemed like a logical place for me to land," he says.

Pharma and CRO work have similarities, but Kremidas says that CROs require employees to be more flexible and nimble: "You have a lot of different customers who have a lot of different ways of doing things." Some customers micromanage, whereas others let the CRO researchers make their own decisions. "They respect your opinions and it's more collaborative," he says. **C.S.**

CONSULTANCY

Use your skills

In many ways, Beat Widler was ideally placed to start a consultancy. He had spent decades in regulatory affairs and clinical research at Roche, the pharmaceutical firm based in Basel, Switzerland. Most recently, he was global head of clinical quality, ensuring that clinical trials protected human subjects and maintained data integrity. Now, he is a consultant in the same area.

Working out of his home in Zug, Switzerland, Widler takes advantage of a network of contacts in the pharmaceutical industry, contract-research organizations and regulatory agencies. Even so, setting up a company was risky. "If we are able to break even this year we can be extremely proud of ourselves," he says.

Widler had been thinking for years about starting a company. When Roche offered him an early-retirement package in 2011, he took the plunge, setting up Widler & Schiemann with former Roche colleague Peter Schiemann this year.

Widler says that the shifts in the industry are making it easier for former pharma employees to set up shop, as big companies and small biotechs turn to an outsourcing model with low overhead costs. And cost pressures are leading companies to rely on experts to help them trim the fat from their clinical trials, while keeping standards high.

Widler's network includes connections at professional organizations such as the European Forum for Good Clinical Practice in Brussels and the Drug Information Association, based in Horsham, Pennsylvania, where he has served on committees and given talks. That experience, he says, helped him to build his reputation and meet clients.

Without the infrastructure of a large organization, Widler has had to adapt. For example, he spent hours creating the template for a form for auditing a client. "You do everything from scratch. It's pretty intense, but it's pretty fun," he says.

He has no regrets and remains optimistic, but is mindful of how long his personal funds can last while he builds up his business. "Be very realistic about finances," he says. "It is critical to do your homework." **C.S.**

TURNING POINT
Jim Hoch

In April, Jim Hoch, a molecular biologist at the Scripps Research Institute in San Diego, California, celebrated the ninth renewal of the grant supporting his study of bacterial signalling proteins. Here, he reflects on how his efforts to unravel sporulation led to a three-decade US National Institutes of Health (NIH) grant — one of the longest-running at Scripps.

How did your research get started?
I came to Scripps after an NIH-funded postdoc at the Molecular Genetics Centre in Gif-sur-Yvette, France, where I learned about mapping genetics in bacteria. Once here, I applied the technique to begin to sort out the mechanisms that trigger sporulation, the process by which bacteria or fungi suspend their growth to form tough, seed-like spores. Since we started, my team and I have learned about the genes and proteins involved, but we are still piecing together how it works. The signalling mechanisms are a mess to unravel.

Was your first grant, to study the bacterium *Bacillus subtilis*, a turning point in your career?
I owe my career to this grant from the US National Institute of General Medical Science. I've been told that you should have several grants, but I had just one that I bundled everything into. If I had lost that grant at any renewal, I would have been dead. The renewal process is fairly traumatic, but it has motivated me to work hard each grant cycle. Because Scripps started hosting graduate students only recently, I have employed technicians, various undergraduates and some postdocs. And we managed to do pretty amazing things.

How have modern technologies influenced your research?
The evolution of technology has driven the experiments. I started with genetics, but molecular cloning and DNA sequencing changed everything — letting us find out about the proteins encoded by the genes. Using biochemistry, we could work out their functions. From there we used crystallography and nuclear magnetic resonance to establish the structure of the proteins. Most recently, we have been working with statistical physicists to determine how these proteins interact. Every year is a complete learning process — a quantum leap from one technology to the next. It has been a hell of a journey.

How have NIH requirements changed?
Proposals used to be more than 20 pages long, and the study sections that review grants lasted for three days. Now, proposals are 12 pages, study sections last one day and half of the applications, the less-impressive ones, are not even discussed. There is a lesson here. Applications of 12 pages need to be clear and concise to make them understandable outside the field. Most importantly, they need to be exciting to read. A proposal needs to have clearly articulated goals that transmit your excitement.

Have you ever thought that your grant wouldn't get renewed?
It has become more difficult over the years; there is more competition. I was worried in the latest round. I broke my leg and was recovering for a year. Normally, I have five or six publications a year, so that every four years, when it is time for renewal, I have 20–25 papers showing my progress. I didn't have that this time, but I squeaked by with a few good papers in good journals.

Your career has been mostly in basic research. Have there been any interesting applications?
There have been some spin-offs. When I first came to Scripps, I was working on genes involved in the hyperproduction of proteolytic enzymes. One of my postdocs ended up as an executive at the biotechnology company Genentech, based in South San Francisco, California. He recognized that proteases could be important in the production of detergent. The proteases in most US soaps come from different species of *Bacillus*, and from some hyperproduction genes that we discovered. A company was spun off: Genencor, which is now owned by Dupont. It has produced more than US$1 billion's worth of enzymes. ■

INTERVIEW BY VIRGINIA GEWIN